

日常の手指動作で人工物操作を行うための手指形状 推定システム

著者	富田 元將
発行年	2016
学位授与大学	筑波大学 (University of Tsukuba)
学位授与年度	2015
報告番号	12102甲第7720号
URL	http://hdl.handle.net/2241/00143871

日常の手指動作で人工物操作を行うための
手指形状推定システム

2016年 3月

富田 元將

日常の手指動作で人工物操作を行うための
手指形状推定システム

富田 元將

システム情報工学研究科
筑波大学

2016年 3月

概要

近年,自由に動きながらのコンピュータグラフィックス制御やロボット操作のような,ユーザの身体動作,特に日常の手指動作で人工物を操作できる入力システムを必要とする多様な人工システムが登場している.しかし,このような入力システムに適したユーザへの負担が少ない手法として,様々なカメラによる3次元手指形状推定手法が提案されているが,データグローブのように十分な分解能を持った手指形状推定を高速高精度に実現する手法はほとんど存在しない.従来研究の中では,著者が知る限りでは唯一,K.Hoshino, et al.によって提案された,事前に特徴量空間における類似度に従って並べ替えたデータベースを準備し,過去の推定結果からデータベースの探索範囲を狭める手指形状推定手法が,入力システムに適した手指形状推定を実現していた.しかし,この手法も,特徴量空間での類似度による並べ替えを,実際の手指形状の類似度に最適化することが難しく,適切にデータベースの探索範囲を狭めることは困難であった.そのため,手指動作によっては,本来推定されるべき真値となるデータが探索範囲外に存在するという問題が発生してしまう.その場合,真値となるデータを出力することはできず,データベース探索範囲内の最も類似する別のデータを出力してしまうため,推定誤差が大きくなってしまう.さらに,高速化のためにデータベースの探索範囲が固定長であるので,一度でも真値となるデータが探索範囲外に存在してしまった場合,基本的には,それ以降の推定でも真値が探索範囲外に存在することになり,推定誤差の大きい状態が続くという問題があった.さらに,ロボット操作では,把持動作に必要な4指と拇指の対向動作を推定できる必要があるが,最も自己遮蔽の影響が強い姿勢であるために,正確な推定は難しい.その上,モーションキャプチャーなどと組み合わせユーザが自由に動作する環境での推定では,カメラ位置を固定しているために,手指が極端に小さく撮像されたり,自身の体で遮蔽してしまうといった問題もあり,入力システムとして利用するためには様々な問題を抱えていた.

そこで,本研究では,全人工システムを,自由に動きながら,物理世界と同じ手指動作で操作するための手指形状推定システムの実現を目指す.実現のために,第1の要素として,過去の推定結果に影響されずに,データグローブのような十分な分解能を持った高速高精度の手指形状推定できる必要がある.また,第2の要素として,把持動作のように,特に自己遮蔽の影響によって推定が難しい4指と拇指の対向動作であっても,高精度に推定できる必要がある.第3の要素として,モーションキャプチャのようにユーザが自由に動作する状況下であっても,高速高精度に推定できる必要がある.上記の3つの要素を満たすことによって,全人工システムを操作可能な手指形状推定システムを実現することができる.

第1の研究では,データベース探索範囲を狭めない手法によって,コンピュータグラ

フィックス制御のための手指形状推定システムの実現を目指した。データベースの探索範囲を狭めない手法として、大まかな形状を表す画像形状比率によるスクリーニングと詳細な形状を表す画像特徴量による類似度計算による2段階のデータベース探索手法を提案した。この手法であれば、データベースを高速に全探索することができるため、探索空間を狭める検討をする必要がない。データグローブの計測値を真値として推定精度を検証するために、データグローブを装着した状態で手指形状推定を行い、データグローブの実測値と推定値を同時に取得して評価実験を行った。その結果、推定誤差の平均値 -2.11 度、標準偏差 ± 14.14 度、処理速度 80fps 程度と高速高精度に実現することができた。

第2の研究では、ロボットの把持動作で最も必要となる4指と拇指の対向動作を高精度に推定するための手指形状推定システムの実現を目指した。4指と拇指の対向動作は、他の手指形状よりも自己遮蔽の影響の大きく、特に、1台のカメラを用いた手指形状推定システムでは、同じ手指のシルエット画像であっても手指形状が異なるために、4指と拇指の対向動作を高精度に推定することは難しい。そこで、本章では、直交関係にある2台のカメラを使い、撮像されている手指面積の大きい方のカメラから4指の手指形状、小さい方のカメラで拇指対向の手指形状を推定するシステムを提案する。この手法により、1台では死角となってしまう部分についても高精度に推定することができる。また、通常、複数のカメラを使う場合には、カメラ同士の厳密な相対位置を検出しておく必要があるが、本手法では厳密な相対位置の検出の必要のない推定手法を使うことで、簡単にシステムを設置することができるシステム構築を実現した。評価実験では、右手にデータグローブを装着した状態で左手で推定を行い、右手と左手を注意深く同期して動かすことで、真値となるデータグローブの実測値と推定値を同時に取得して評価実験を行った。その結果、推定誤差の平均値と標準偏差は示指 PIP 関節で 0.45 ± 14.57 度、拇指 CM 関節で 4.7 ± 10.82 度であり、最も重要となる4指と拇指の対向動作についても正確に推定できていることが確認できた。また、実際にロボット操作へ応用したシステムでは、ロボットを通じた物体把持動作によって、安定して対象の人工物を把持することができた。

第3の研究では、第1、第2の研究のようにユーザがほとんど動きまわらないシステムと違い、モーションキャプチャシステムのように、広い空間で自由に動き回り、様々な姿勢をとるユーザを対象とする状況に対応できる手指形状推定システムの構築を目指した。2章と3章で示した方法は、カメラ位置が空間に対して固定されているため、撮像される手指が非常に小さくなったり、体で隠されてしまったりするという問題があった。そこで、超小型カメラと小型コンピュータによる装着型ハードウェアを使った手指形状推定システムによって、オペレータの動きを妨げずに、手指形状推定できるシステムの実現を図った。超小型カメラは、小さいもので $7\text{mm} \times 7\text{mm}$ 程度と非常に小さいので、装着してもオペレータへの負担が少なく、自由な動きを妨げない。また、試

作機では、小型カメラを手の甲から約 7cm 程度上の部分となるように装着しており、ユーザの動作への影響を最小限に抑えている。ただし、小型コンピュータは通常の PC よりもスペックが低いため、高速に動作する手指形状推定手法が必要となる。そこで、本章では、画像特徴量の算出に使用していた HLAC を位置情報のみ使用する方法で、1600 次元から 64 次元に次元圧縮する手法を提案した。この手法により、従来と同程度の推定精度を維持しつつ、高速に推定することができるようになった。評価実験は、第 2 の研究と同様に、右手にデータグローブ、左手に小型カメラと小型コンピュータを組み合わせたハードウェアを組み合わせた試作機を装着して実施した。その結果、推定誤差の平均値と標準偏差は -5.7 ± 7.2 度であり、推定速度の平均値は 25.0ms であった。一方、1600 次元の画像特徴量を使う手法では、推定精度が -6.0 ± 6.9 度で、推定速度の平均値が 158.7ms であり、ほとんど推定精度を落とすことなく、推定速度を大幅に高速化することに成功した。

第 1, 第 2, 第 3 の研究の結果と、それらの総合的考察から、本研究が最終目標とする物理世界と同じ手指動作で、自由に動きながら、コンピュータ・グラフィックス制御とロボット操作を可能にする手指形状推定システムが達成されたと結論付けることができる。

目次

1 章 序論.....	1
1.1 研究の背景.....	1
1.2 目的.....	3
2 章 関連研究.....	5
2.1 手指形状推定の概要.....	5
2.2 3D-model-based アプローチ.....	5
2.3 2D-appearance-based アプローチ.....	6
3 章 2 段階のデータベース探索による手指形状推定手法.....	8
3.1 はじめに.....	8
3.2 システム構成.....	9
3.2.1 データベース構造.....	9
3.2.2 基準点と手指範囲の検出.....	11
3.2.3 画像形状比率の算出.....	12
3.2.4 画像特徴量の算出.....	13
3.2.5 前腕傾きの算出.....	14
3.2.6 2 段階のデータベース探索.....	16
3.3 評価.....	17
3.3.1 画像形状比率によるスクリーニングの閾値決定のための予備実験.....	17
3.3.2 評価実験.....	20
3.4 まとめ.....	26
4 章 ゆるい直交関係にある 2 台のカメラを使った手指形状推定システム.....	27
4.1 はじめに.....	27
4.2 システム構成.....	29
4.2.1 2 視点からの画像情報を持つデータベース構造.....	29
4.2.2 推定手法.....	30
4.3 評価.....	31
4.4 まとめ.....	36
5 章 超小型カメラを使ったウェアラブル手指形状推定システム.....	37
5.1 はじめに.....	37
5.2 システム構成.....	39
5.2.1 ウェアラブル入力デバイス.....	39
5.2.2 画像特徴量の次元圧縮.....	40
5.2.3 データベース構造.....	41
5.2.4 推定手法.....	42
5.3 評価.....	43

5.3.1 次元圧縮検証のための予備実験.....	43
5.3.2 評価実験	45
5.4 まとめ	49
6 章 結論.....	50
謝辞	52
参考文献	53
研究業績	56

図目次

図 3.1 データベース構造.....	10
図 3.2 基準点と手指範囲の検出.....	11
図 3.3 画像形状比率の算出.....	12
図 3.4 画像特徴量の算出.....	13
図 3.5 前腕傾きの補正	14
図 3.6 手指, 前腕領域輪郭線の各画素周りに対する接線の傾き	15
図 3.7 手指, 前腕領域輪郭線の各画素周りに対する接線傾きの標準偏差.....	15
図 3.8 各閾値を変えた時の画像形状比率によるスクリーニングの平均通過数.....	18
図 3.9 3 つの閾値を同時に変化させた時の画像形状比率によるスクリーニングを通過する平均データセット数と推定失敗確率の関係	19
図 3.10 評価実験の撮像条件 (左) とカメラから取得できる映像 (右)	20
図 3.11 CG による手指形状推定の描画.	23
図 3.12 データベース作成者の手による形状推定結果例	24
図 3.13 別の被験者の手による形状推定結果例	25
図 4.1 自己遮蔽により拇指形状の違いが分からない例	28
図 4.2 2 視点からのカメラ構成概要	29
図 4.3 2 視点からの情報を持つデータベース構造	29
図 4.4 2 視点からの画像情報を使ったデータベース探索	30
図 4.5 評価実験環境の概要	31
図 4.6 推定結果のスナップショット	33
図 4.7 データグローブによる推定結果例	34
図 4.8 手指形状推定を用いた遠隔ロボット操作による安定した物体操作.....	35
図 5.1 ウェアラブルデバイスの構成概要	39
図 5.2 画像特徴量の次元圧縮概要	40
図 5.3 次元圧縮した画像特徴量を使ったデータベース構造.....	41
図 5.4 カメラキャリブレーション	42
図 5.5 画像特徴量の次元の違いによる推定結果	44
図 5.6 評価で使用した試作機	45
図 5.7 CG による推定結果例 (各写真の右が入力の手指形状, 左が推定結果の CG)	47
図 5.8 推定結果例	48
表 5.1 画像特徴量の次元数の違いによる推定誤差	43

1 章 序論

1.1 研究の背景

近年, ユーザの身体動作を使って人工物を操作できる入力システムを必要とする多様なシステムが登場している. 例えば, 3D のコンピュータグラフィックス制御や複雑な多関節構造を持つロボットなどの操作は, マウスやキーボードといった既存の入力システムでは操作が非常に難しく, 操作できるといっても限定的な操作しかできない. これらのシステムを自在に操作するために, 新しい入力システムが求められている.

最も簡単に操作するための入力システムとして, 日常の手指動作をそのまま入力として使う手法が考えられる. 日常生活において, ヒトは自在にモノを扱うことができ, 特に, 最も器用に動かすことができる手指を使えば, 誰でも簡単に, 高度で複雑な操作をすることができる. つまり, ヒトの手指動作をそのまま入力として使うことのできる入力システムがあれば, 誰でも簡単にコンピュータグラフィックスやロボットなどの人工物操作を実現することができる.

既存の手指動作を取得するシステムは, 拘束によるユーザへの負荷が大きかったり, 手指動作を検出可能な範囲が狭いなどの問題があり, あらゆる状況に対応するための入力システムとしては不向きであった. 例えば, 代表的な入力システムの 1 つであるデータグローブは, グローブを装着した使用者が手指を動かすと, 内蔵されているひずみゲージの曲がり具合から指関節の角度を計測することができる. しかし, このグローブは, 手に密着させる必要があるために締め付けがきついためユーザの負担が大きく, また機材が非常に高価であるため, 入力システムとして使用することは難しい. また, その他の方法としてデプスカメラを使う手法[1]-[2]も提案されているが, GPU を使用する必要があるほど計算コストがかかり, また, 解像度が低く検出可能な範囲が狭く, ユーザが自由に動くことができないため, 入力システムに適していない.

ユーザへの負担が少なく, 動作を妨げない手法として, カメラを使った 3 次元手指形状推定手法がある. この手法では, ユーザがカメラの前で手指を動かし, 画像認識によって各指関節角度として手指動作を取得する方法である. この手法であれば, 非接触であるためユーザの手指動作に負担をかけることはなく, 撮像範囲も非常に広いのでユーザの自由な動きを妨げることもない. カメラによる 3 次元手指形状推定システムは, 入力システムとして利用するのに, 最も有効な手法であると考えられる.

3 次元手指形状推定手法は, 大きく 3D-model-based アプローチ[3]-[11]と 2D-appearance-based アプローチ[12]-[25]の 2 種類に分けることができる. 3D-model-based アプローチは, カメラで撮影した入力画像中のシルエットのような局所的な特徴に対して, 事前に計算機内部に構築しておいた 3 次元手指モデルを当てはめる手法である. この手法は, 原理的に高精度の推定が可能であるものの, 長い処理時間が必要となるため, リアルタイムでの処理が必要とされる入力システムには適していない. 一方,

2D-appearance-based アプローチは、事前に構築した様々な手指形状データを持つデータベースから、入力画像から抽出した特徴を使ってデータベースを探索し、再類似の手指形状データを出力する手法である。この手法で高精度な推定を実現するためには、様々な種類の手指形状データを増やし、データベース規模を大きくする必要がある。そのため、高精度な推定のためには、大規模なデータベースからの探索が必要となるが、処理内容が比較的単純であり、計算時間の短縮が原理的には可能であるので、リアルタイムで推定できる必要のある入力システムに適している。ただし、既存の 2D-appearance-based アプローチの手法のほとんどは、数十～数百種類の手指形状パターンへの分類程度しか実現できず、データグローブのように十分な分解能を持つ推定を実現することができない。これでは、手指のジェスチャーによる単純なボタン操作や手話の認識といった目的程度であれば利用できるものの、本研究が目的とするコンピュータグラフィックス制御やロボット操作を実現することはできない。本研究の目的を達成するためには、少なくとも数千～数万種類の手指形状パターンから手指形状推定を推定し、データグローブのように十分な分解能を持って関節角度を推定できる必要がある。

データグローブのように十分な分解能を持つ高速高精度な手指形状推定は、Hoshino, et al. [19]-[25]によって提案されており、事前に特徴量空間における類似度に従って並べ替えたデータベースを準備し、過去の推定結果からデータベースの探索範囲を狭める手法を提案している。この手法は、データグローブのような十分な分解能を実現するための数千種類の手指形状パターンを持つデータベースを使った高速高精度な手指形状推定を実現している。入力システムに適した手指形状推定を実現していた。しかし、この手法は、特徴量空間での類似度による並べ替えを、実際の手指形状の類似度に最適化することが難しく、適切にデータベースの探索範囲を狭めることは困難であった。そのため、手指動作によっては、本来推定されるべき真値となるデータが探索範囲外に存在するという問題が発生してしまう。その場合、真値となるデータを出力することはできず、データベース探索範囲内の最も類似する別のデータを出力してしまうため、推定誤差が大きくなってしまう。さらに、高速化のためにデータベースの探索範囲が固定長であるので、一度でも真値となるデータが探索範囲外に存在してしまった場合、基本的には、それ以降の推定でも真値が探索範囲外に存在することになり、推定誤差の大きい状態が続くという問題があった。また、ロボット制御や自由な動作をするユーザに対する推定を実現するためには、ソフトウェアのみでなくハードウェア構成を含めた検討が必要になる。ロボット操作では、最も期待されるロボットを通した物体把持が実現のために4指と拇指の対向動作を正確に推定する必要があるが、自己遮蔽の問題を解決する必要がある。特に、把持動作では指同士が最も重なる状態での手指形状の違いを推定できる必要がある。さらに、このようなユーザが動きまわらないシステムと違い、モーションキャプチャと併用するような、自由に動き回るユーザに対しても手指形状推定できる必要がある。この場合、ユーザは自由に動きまわるので、カメラ位置を固定してしま

うと、撮像される手指のサイズが極端に小さくなったり、自身の体で遮蔽してしまう問題が発生してしまっていた。

1.2 目的

本稿では、全人工システムを、自由に動きながら、物理世界と同じ手指動作で操作するための手指形状推定システムの実現を目指す。実現のために、第1の要素として、過去の推定結果に影響されずに、データグローブのような十分な分解能を持った高速高精度の手指形状推定できる必要がある。また、第2の要素として、把持動作のように、特に自己遮蔽の影響によって推定が難しい4指と拇指の対向動作であっても、高精度に推定できる必要がある。第3の要素として、モーションキャプチャのようにユーザが自由に動作する状況下であっても、高速高精度に推定できる必要がある。上記の3つの要素を満たすことによって、全人工システムを操作可能な手指形状推定システムを実現することができる。

まず、第1の要素を実現するために、第3章では、手指画像から低次の粗い画像形状比率と高次の精細な画像特徴量を算出し、第1段階を画像形状比率によるスクリーニング、第2段階を画像特徴量による類似度計算とする2段階のデータベース探索手法を採用することで、データベースの探索範囲を狭める必要のない手指形状推定を提案する。この手法は、データベースの探索範囲を狭める必要のない手法であるので、狭めることによって発生する問題を解決することができる。

次に、第2の要素を実現するために、第4章では、ロボットの把持動作で最も必要となる4指と拇指の対向動作を高精度に推定するための手指形状推定システムの実現を目指す。ロボットによる把持動作を実現するためには、4指と拇指の対向動作を高精度に推定できることが欠かせない。しかし、把持動作は、他の手指形状よりも自己遮蔽の影響の大きく推定が難しい。特に、1台のカメラを用いた手指形状推定システムでは、同じ手指のシルエット画像であっても手指形状が違うために、4指と拇指の対向動作を高精度に推定することは非常に難しい。そこで、直交関係にある2台のカメラを使い、撮像されている手指面積の大きい方のカメラから4指の手指形状、小さい方のカメラで拇指対向の手指形状を推定するシステムを提案する。この手法により、1台では死角となってしまう部分についても高精度に推定することができる手指形状推定システムの実現を目指す。

最後に、第3の要素を達成するために、第5章では、第3章と第4章のシステムのように、ユーザがほとんど動かないシステムではなく、モーションキャプチャシステムのように、広い空間で自由に動きまわり、様々な姿勢をとるユーザを対象とする状況に対応できる手指形状推定システムの構築を目指した。上述の手法は、カメラ位置が空間に対して固定されているため、撮像される手指が非常に小さくなったり、体で隠されてしまったりするという問題があった。そこで、超小型カメラと小型コンピュータによる

装着型ハードウェアを使った手指形状推定システムによって、オペレータの動きを妨げずに、手指形状推定できるシステムの実現を図った。ただし、小型コンピュータは通常の PC よりもスペックが低いため、高速に動作する手指形状推定手法の開発も必要不可欠となる。そこで、画像特徴量を 1600 次元から 64 次元に次元圧縮する手法を提案する。この手法により、自由に動作するユーザであっても高精度に推定可能な手指形状推定システムの実現を目指した。

2 章 関連研究

2.1 手指形状推定の概要

コンピュータビジョンによる手指の認識技術は、手話のように数少ない手指パターンを判別するジェスチャー認識技術と、各指関節角度を詳細に推定する手指形状推定技術に大別することができるが、本研究では後者を対象としている。手指形状推定では、対象が非常に複雑な多関節構造であり、さらに自己遮蔽の影響が極めて大きいため、難解な問題のひとつとして研究が進められている。

手指形状推定では、20 もの自由度を持つ多関節構造にある手指を推定の対象としている。そのため、推定の分化能を細かくすればするほど、計算コストが指数関数的に増大するという問題も発生してしまう。これは、単純に考えると、各指関節角度の分解能を上げると、関節の自由度である 20 乗で計算コストに跳ね返ってくるためである。しかし、手指の動きは非常に早く、およそ $300^{\circ}/s$ とも言われており、リアルタイムでの推定を実現するためには、少なくとも 30~60Hz 必要と言われている。

また、この複雑さゆえに、手指は自己遮蔽しやすい構造となってしまう。一般に、コンピュータビジョンでは、自己遮蔽している部分が増えると、画像の特徴の変化を検出することが難しくなり、原理的に推定が困難になってしまう。そのため、手指形状推定では、指同士や掌と指の重なりといった自己遮蔽に対応する必要がある。

このように、入力システムに必要となる手指形状推定手法には、自己遮蔽の多い多関節構造物である手指から高い推定精度を実現する必要がある上に、高い推定精度を維持しつつも高速な推定速度を両立する高度なアルゴリズムが必要となる。一般に、手指形状推定手法は、大きく 3D-model-based アプローチと 2D-appearance-based アプローチの 2 種類に分類することができ、以下に紹介するように、様々な手法が提案されている。

2.2 3D-model-based アプローチ

3D-model-based アプローチの手指形状推定手法は、システムが持つ 3 次元の手指モデルを使って手指形状を推定する手法である。この方法は、入力画像から抽出される特徴に対して、3 次元の手指モデルを当てはめることによって手指形状の推定を行う。一般的な性質として、手指モデルを使うことで高精度の推定が実現できる一方で、計算コストが高く、推定速度が遅いという特徴がある。

例えば、S.Lu, et al.[3]は、掌を六角柱、指の各部位を円柱として連結させたシンプルな手指モデルを使って手指形状推定を行っている。この手法では、まず、システムが持つ手指モデルの手指形状と入力となる手指の姿勢が一致するようにキャリブレーションし、オプティカルフローの情報を元に、毎フレーム入力画像の手指の移動量を算出し、手指モデルの手指形状を更新することで、推定を行っている。ただし、高精度の推定は実現できているものの、推定速度は 4fps 程度でしかない。そのため、オプティカルフ

ローを使っているこの手法では、1 フレームでの変化量が極端に大きいと正確に推定することができないために、入力システムとして使う場合、ユーザは手指を十分にゆっくり動かす必要があり、入力システムに向いているとは言えない。他にも、Y.Wu, et al.[4]によってモンテカルロ法をベースとした手指形状推定手法や、島田ら[5]による複数のカルマンフィルタと分布切取り法を用いた手法、亀田ら[6]によるトップダウンに手指モデルに投影したシルエットに対する被覆領域からの推定手法、E.Ueda, et al.[11]の手を囲むように配置した複数のカメラ画像から作成した Voxel モデルと手指モデルをマッチングする手法や、T.Gumpp, et al.[4]の人間の眼幅と同じ距離に配置した 2 台のカメラから取得した視差画像を元に作成したデプス画像と手指モデルから推定する手法など、様々な推定手法が提案されているが、いずれもリアルタイムで推定を実現できたとは言えず、入力システムとして使うためには十分な性能を持っていない。

2.3 2D-appearance-based アプローチ

2D-appearance-based アプローチによる手指形状推定手法は、システムが様々な手指形状に関するデータを保持したデータベースを使う手法である。この方法は、入力画像から抽出去れる特徴と各手指形状のデータが持つ特徴を照合し、最類似の手指形状データを抽出することで手指形状を推定する手法である。一般的な性質として、データベースの探索手法を最適化することで高速な推定ができる一方で、ほとんどの研究は、数十から数百程度のパターン分けが実現できる程度である。そのため、データグローブのように十分な分解能を実現しているものは少なく、ほとんどの研究では、パターンへの分類の正答率によって推定精度を検証している。

例えば、Athitsos, et al.[12]は、セグメンテーション分けした手指モデルを元に、様々な視点の画像から抽出した輪郭線の特徴量とする手法で手指形状推定を実現している。この手法では、推定速度 15fps とある程度のリアルタイム性を実現しているものの、推定できるパターンは、26 パターン程度であり、十分な分解能を持った推定ができると言えない。また、他の研究では、Ohkawa, et al.[13]による非線形カーネル直交相互部分空間法(KOMSM)を使った手法や、M.H.Jeong, et al.[14]による EM アルゴリズムで学習し coupled-forward アルゴリズムを使った手法、Yasumoto[15]の K-means 法とガウシアンカーネル関数を組み合わせた手法など、様々な手法が提案されているが、いずれもパターン数が非常に少なく、指文字や手話、ジェスチャー認識などのパターン数の少ないシステムへの応用は可能であるものの、本研究が目的とする手指動作の入力システムとして利用することできなかった。

逆に、多くのデータベースから推定する方法としては、今井ら[16]が、3 次元の手指モデルから生成した CG を使って、事前に典型的な見え方に対して複数のクラスタを生成し、推定する手法も提案されている。しかし、5,000 パターンからの推定を実現するものの、1 フレーム当たり 3.6 秒の処理時間も必要としてしまうため、入力システムの

ようなリアルタイムでの処理を実現することは難しかった。

大規模なデータベースから高速に推定できる手法としては、Hoshino, et al[19]-[25]が数千のデータベースから 160fps と高速な手指形状推定手法を実現していた。この手法では、事前に特徴量空間における類似度に従って並べ替えたデータベースを準備し、過去の推定結果からデータベースの探索範囲を狭める手法を提案している。しかし、特徴量空間での類似度による並べ替えを、実際の手指形状の類似度に最適化することが難しく、適切にデータベースの探索範囲を狭めることは困難であった。仮に、先験的知識を用いてデータベースを多階層化しようとした場合でも、代表的な値を持つ各階層の各クラスに、画像特徴に関するどのような情報を割り当てたら良いかを決めるのが容易でなかった。そのため、手指動作によっては、本来推定されるべき真値となるデータが探索範囲外に存在するという問題が発生してしまう。その場合、真値となるデータを出力することはできず、データベース探索範囲内の最も類似するデータを出力してしまうため、推定誤差が大きくなってしまう。さらに、高速化のためにデータベースの探索範囲が固定長であるので、一度でも真値となるデータが探索範囲外に存在してしまった場合、基本的には、それ以降の推定でも真値が探索範囲外に存在することになり、推定誤差の大きい状態が続くという問題があった。

3 章 2 段階のデータベース探索による手指形状推定手法

3.1 はじめに

従来手法[19]–[25]では、事前に特徴量空間における類似度に従って並べ替えたデータベースを準備し、過去の推定結果からデータベースの探索範囲を狭める手法を提案している。この手法によって、数千種類の手指形状パターンを持つデータベースを使った高速高精度な手指形状推定を実現している。しかし、特徴量空間での類似度による並べ替えを、実際の手指形状の類似度に最適化することが難しく、適切にデータベースの探索範囲を狭めることは困難であった。仮に、先験的知識を用いてデータベースを多階層化しようとした場合でも、代表的な値を持つ各階層の各クラスに、画像特徴に関するどのような情報を割り当てたら良いかを決めるのが容易でなかった。そのため、手指動作によっては、本来推定されるべき真値となるデータが探索範囲外に存在するという問題が発生してしまう。その場合、真値となるデータを出力することはできず、データベース探索範囲内の最も類似するデータを出力してしまうため、推定誤差が大きくなってしまう。さらに、高速化のためにデータベースの探索範囲が固定長であるので、一度でも真値となるデータが探索範囲外に存在してしまった場合、基本的には、それ以降の推定でも真値が探索範囲外に存在することになり、推定誤差の大きい状態が続くという問題があった。

そこで、本章では、手指画像から低次の粗い画像形状比率と高次の精細な画像特徴量を算出し、第1段階を画像形状比率によるスクリーニング、第2段階を画像特徴量による類似度計算とする2段階のデータベース探索手法を採用することで、データベースの探索範囲を狭める必要のない手指形状推定を提案する[26]。画像形状比率は、大まかに前腕回旋と4指屈曲伸展、拇指屈曲伸展を反映する3つの比率情報であり、この比率でデータベースを高速にスクリーニングすることで推定の高速化を図る。また、画像特徴量は、N.Otsuら[27]によって提案された高次局所自己相関関数(以下、HLACと称す。)を使った1600次元の特徴量で手指形状を表し、最類似のデータを決定する類似度計算に使用し、高精度な推定を実現する。この方法は、データベースを全探索する手法であるので、探索範囲を狭めることによって、最適なデータが探索範囲外に設定されてしまう問題は発生しない。

3.2 システム構成

3.2.1 データベース構造

本システムのデータベースは、図 3.1 のように、様々な手指形状を表すデータセットの集合であり、各データセットは、画像形状比率と画像特徴量、角度情報の 3 種類の情報で構成されている。画像形状比率は大まかな手指形状を、画像特徴量は詳細な手指形状を表した値であり、後述の画像処理によって入力画像から算出され、データベース探索に使用される。

各データセットの保持する角度情報は、手指形状の実測値を表し、各指関節と手首関節、前腕回旋の角度で構成されている。各指と手首関節は、Virtual Technologies 社製のデータグローブによって取得する。データグローブは 18 の自由度を持ち 24 種の角度データを出力する。このうち、示指、中指、薬指、小指の各 MP 関節、PIP 関節、DIP 関節と拇指 CM 関節、MP 関節、IP 関節、さらに各指間の開閉関節角度の計 19 の関節角度を取得し、データベースに保存する。次に、前腕回旋角度は、掌がカメラに向いている状態を 0 度とし、手小指側面がカメラに向いた時を 90 度、手の甲がカメラに向いた時を 180 度となるように前腕の回旋角度を定めた。回旋角度の取得方法は、手上部に取り付けたカメラから見えるように小指側の手側面部にマーカを装着し、マーカの位置をカメラによって取得することで回旋角度を取得する。

実際のデータベース作成の際には、データグローブ、マーカを装着した状態でカメラの前で手指を動かすことで、各指関節角度、前腕回旋角度をデータグローブ、マーカから取得し、同時にカメラの入力画像から後述の処理によって算出する画像形状比率、画像特徴量を取得することでデータベースの作成を行う。

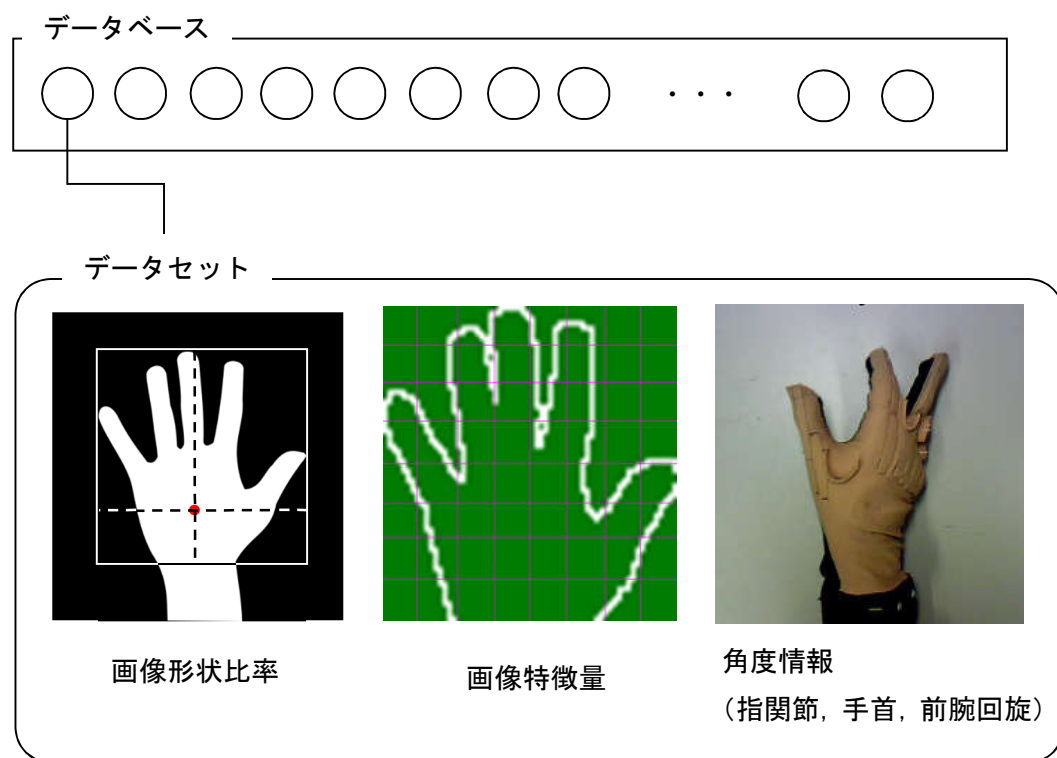


図 3.1 データベース構造

3.2.2 基準点と手指範囲の検出

まず，入力画像から基準点を検出する．基準点の検出では，はじめに，平滑化を施した入力画像から手指と前腕領域と背景領域に 2 値化した画像を作成する．2 値化画像は，膨張縮体処理を施すことで，ノイズを除去しておく．次に，手指と前腕領域の輪郭部分のピクセルをラベル番号 1 として，ラベル番号 i に隣接する手指と前腕領域のピクセルにラベル番号 $i+1$ を与えるラベリング処理を手指と前腕領域全体に施す．その結果，ラベル番号が最も大きい点を基準点として設定する．

次に，基準点から手指範囲の決定を行う．手指範囲は，図 3.2 の手指部分をちょうど囲むような範囲とし，肘から手首へ方向とその法線方向の直線によって囲まれた長方形によって決定される．便宜上，肘から手首へ方向を上向きとする方向を上下方向，上方向から時計回りに 90 度向きを変えた方向を右とする左右方向ということとする．まず，下端部分の決定を行う．下端部分は基準点から最も近い輪郭線までの距離分だけ下方向に下げた場所とする．上端，左端，右端は，下端部分よりも上方向にある手指領域が内接するように決める．こうして決められた上端，下端，左端，右端に囲まれた範囲を手指範囲とする．

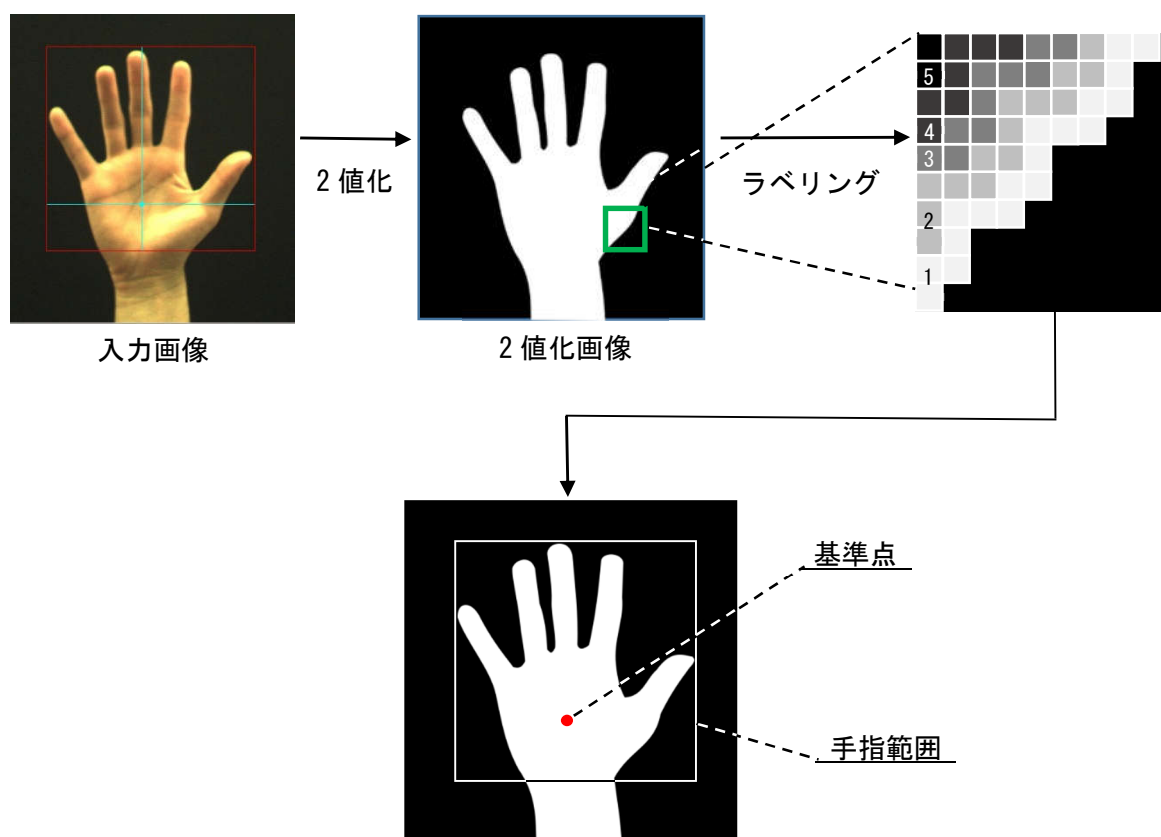


図 3.2 基準点と手指範囲の検出

3.2.3 画像形状比率の算出

画像形状比率は、縦長度と上長度、右長度の 3 つの値で手指の大まかな形状を表す。それぞれ (3.1)～(3.3)式で算出されており、図 3.3 のような基準点と手指範囲と関係している。これらの比率は、縦長度は前腕回旋角度、上長度は 4 指の屈曲伸展、右長度は拇指の屈曲伸展の状態に大まかに対応している。

$$R_{\text{tall}} = H / (H + W) \quad (3.1)$$

$$R_{\text{topheavy}} = H_{\text{upper}} / H \quad (3.2)$$

$$R_{\text{rightbiased}} = W_{\text{right}} / W \quad (3.3)$$

ここで、

R_{tall} ：縦長度

R_{topheavy} ：上長度

$R_{\text{rightbiased}}$ ：右長度

H：手領域の上端から下端までの距離

H_{upper} ：基準点から手指領域の上端までの距離

W：手指領域の左端から右端までの距離

W_{right} ：基準点から手指領域の右端までの距離

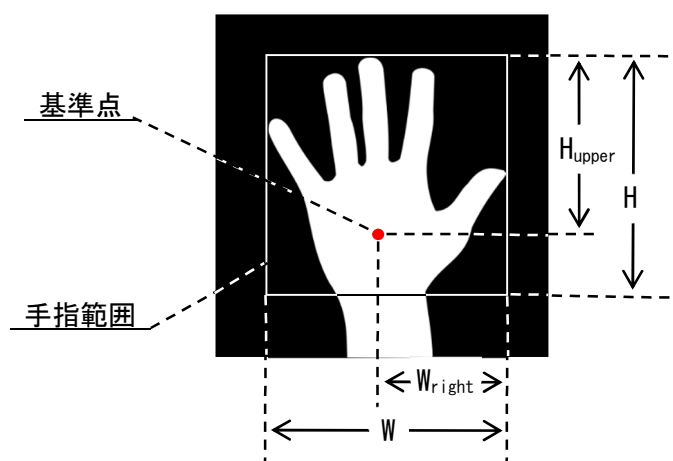


図 3.3 画像形状比率の算出

3.2.4 画像特徴量の算出

詳細な手指画像情報を表す画像特徴量の算出では、図 3.4 のように、正規化した輪郭線画像に対して HLAC を適用し算出する。まず、2 値化画像の手指範囲からシルエットの輪郭線を抽出し、64pixel×64pixel に正規化する。次に、正規化した画像を 8×8 の 64 画面に分割し、各分割画面に対して HLAC を適用することで画像特徴量を算出する。HLAC は各分割画面の輪郭線情報を 25 次元の特徴量で記述することができる。本手法では、分割画面毎に 25 次元の特徴量が算出されるので、1 つの手指形状を 1600 次元の特徴量で表現している。

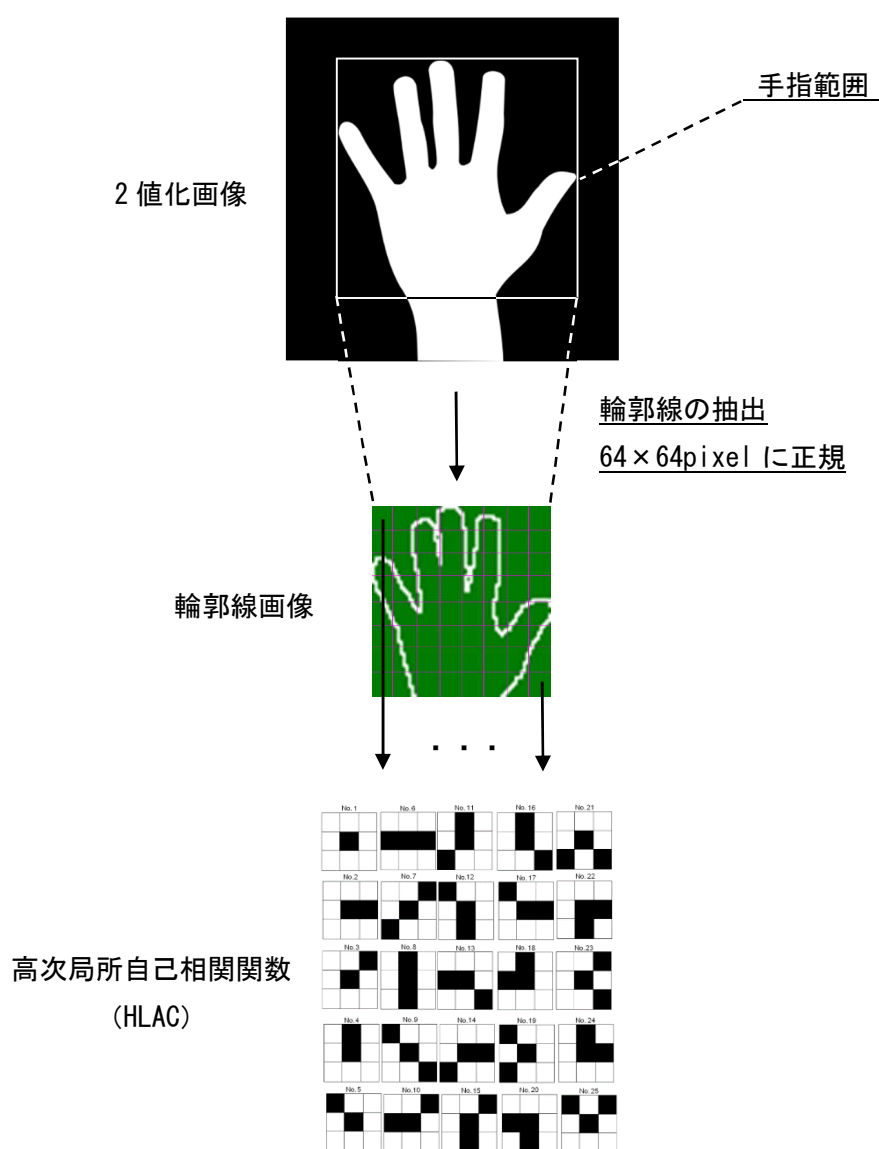


図 3.4 画像特徴量の算出

3.2.5 前腕傾きの算出

推定の際，入力画像の前腕の傾きを算出し，補正する必要がある．まず，図 3.5 に示した 4 点である S ， S' ， E ， E' 点を求める． S ， E 点は画面端にある手指，前腕領域の輪郭線上の画素である． S から E まで輪郭線上の画素を輪郭線に沿って追跡し，各画素の傾きを算出する．各画素の傾きは，対象の画素から前後数 pixel 離れた 2 つの画素を結ぶ直線の傾きとする． S 点から E 点までの傾きをグラフにしたものを図 3.5 に示す．次に，各画素周りの標準偏差を求める．図 3.6 グラフの各画素周り数十 pixel の標準偏差を求めたものを図 3.7 に示す．これにより，傾きの変化の大きい個所では標準偏差が大きくなり，変化の小さい個所は小さくなる．標準偏差がある閾値以上で， S 点に最も近い点を S' 点， E 点に最も近い点を E' 点とする．

S ， S' を結ぶ直線を L_S ， E ， E' 点を結ぶ直線を L_E とする．前腕の傾きは， L_S ， L_E の傾きの平均とする．算出された前腕傾きを元に，図 3.5 右のような前腕傾きを補正した画像を作成する．

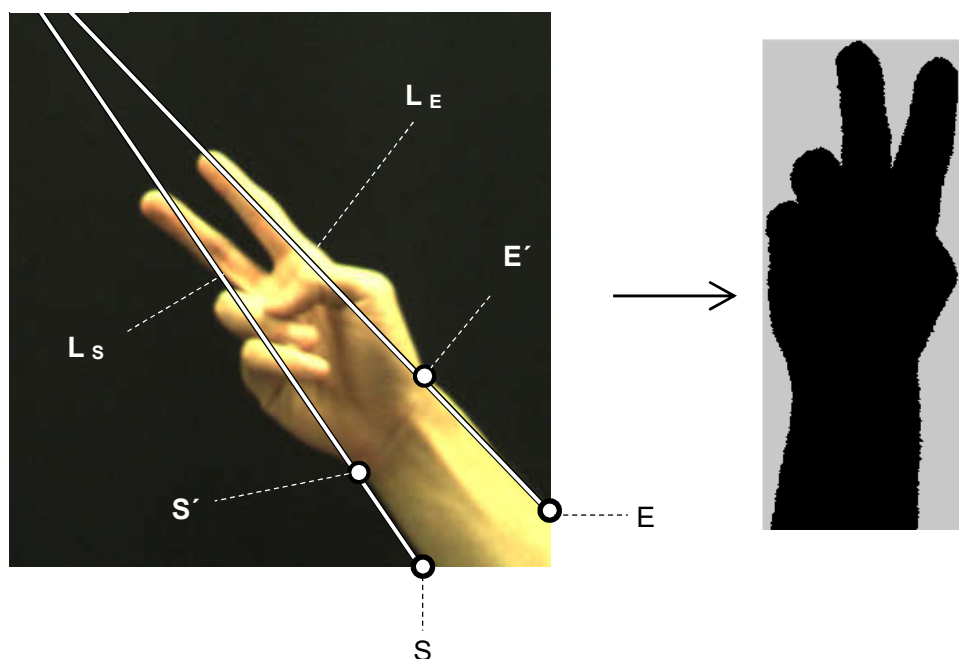


図 3.5 前腕傾きの補正

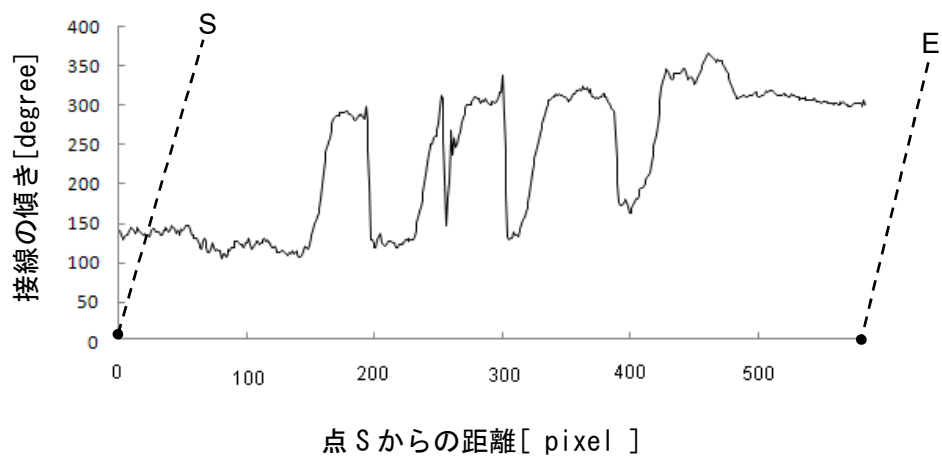


図 3.6 手指，前腕領域輪郭線の各画素周りに対する接線の傾き

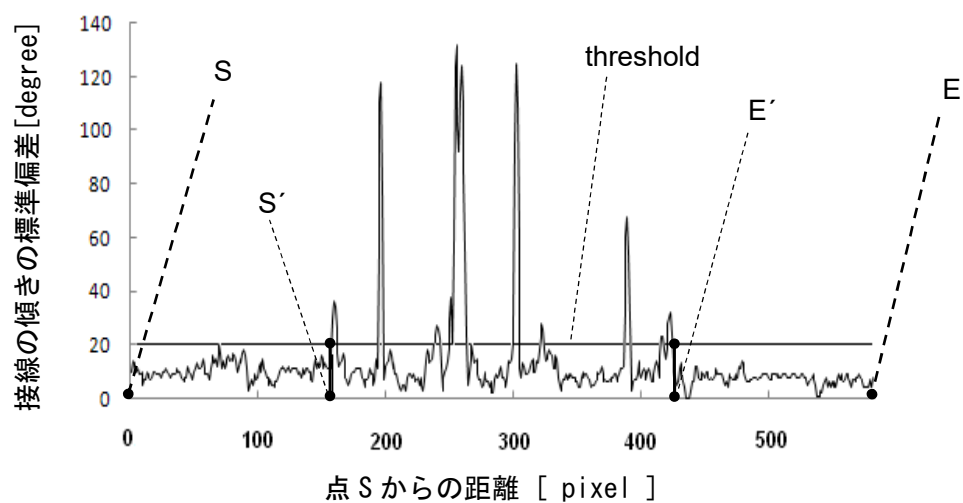


図 3.7 手指，前腕領域輪郭線の各画素周りに対する接線傾きの標準偏差

3.2.6 2段階のデータベース探索

データベース探索では、3.2.5 で取得した前腕と手指のシルエット画像に対して、3.2.2 から 3.2.4 の処理を施して算出した入力画像の画像形状比率と画像特徴量を使用する。

まず、データベースの探索の第 1 次探索として、画像形状比率を使ったスクリーニングを行い、後述の画像特徴量による類似度計算をする対象を絞り込む。スクリーニングでは、全データセットから(3.4)～(3.6)式の全てを満たすデータセットを抽出する。

$$Th_{tall} > |R_{tall}[j] - R_{current-tall}| \quad (3.4)$$

$$Th_{topheavy} > |R_{topheavy}[j] - R_{current-topheavy}| \quad (3.5)$$

$$Th_{rightbiased} > |R_{rightbiased}[j] - R_{current-rightbiased}| \quad (3.6)$$

ここで、

Th_{tall} : 縦長さに関する閾値

$Th_{topheavy}$: 上長さに関する閾値

$Th_{rightbiased}$: 右長さに関する閾値

j : データセット番号

$R_{tall}[j]$: j 番目のデータセットが保持する縦長さ

$R_{topheavy}[j]$: j 番目のデータセットが保持する上長さ

$R_{rightbiased}[j]$: j 番目のデータセットが保持する右長さ

$R_{current-tall}$: 入力画像から取得した縦長さ

$R_{current-topheavy}$: 入力画像から取得した上長さ

$R_{current-rightbiased}$: 入力画像から取得した右長さ

次に、画像特徴量を使って、抽出されたデータセットに対する類似度を算出する。類似度は、(3.7)式で示す単純なユークリッド距離によって算出される。抽出された全てのデータセットと入力画像の類似度を算出し、最も類似するデータセットが持つ角度情報を推定結果として出力する。

$$E[j] = \sum_{l=1}^8 \sum_{c=1}^8 \sum_{h=1}^{25} (x_{current_{lch}} - x_{dataset[j]_{lch}})^2 \quad (3.7)$$

ここで、

j : データセット番号

$E[j]$: 入力画像と j 番目のデータセットとの類似度

l, c : 分割画面番号

h : HLAC パターン番号

$x_{current_{lch}}$: 入力画像の画像特徴量

$x_{dataset[j]_{lch}}$: j 番目のデータセットの保持する画像特徴量

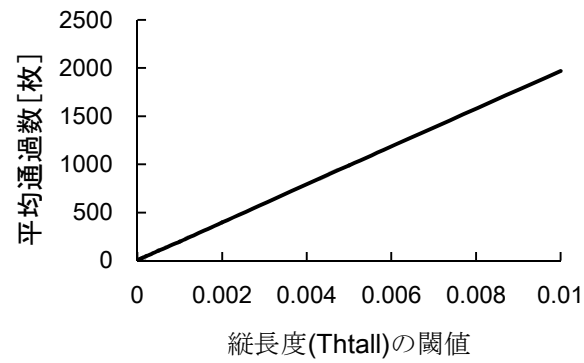
3.3 評価

3.3.1 画像形状比率によるスクリーニングの閾値決定のための予備実験

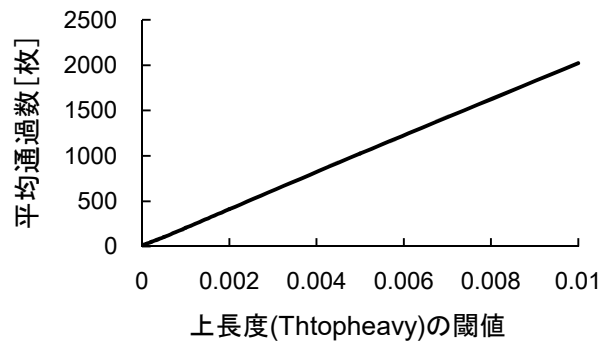
3.2.6 のデータベース探索における画像形状比率によるスクリーニングでは, (3.4)～(3.6)式の各閾値である Th_{tall} , Th_{topheavy} , $Th_{\text{rightbiased}}$ の値を事前に設定しておく必要がある. そこで, 最適な閾値を設定するための予備実験として, 閾値の値に対するスクリーニングを通過するデータセット数と少なくとも 1 セット以上通過する確率の変化を調べた. なお, この実験で使用しているデータベース中のデータセット数は 28,386 セットである.

予備実験の結果を図 3.8 と図 3.9 に示す. 図 3.8 は, 各閾値をそれぞれ変化させた時の画像形状比率によるスクリーニングを通過したデータセット数の関係を示したグラフであり, (a)が縦長さ, (b)が上長さ, (c)が右長さの閾値をそれぞれ変えた時のものである. また, 図 3.9 は, (a)は 3 つの閾値を同時に変化させた時の画像形状比率によるスクリーニングを通過したデータセット数の関係, (b)は, 各推定において 1 セットも通過しない確率を示したグラフである.

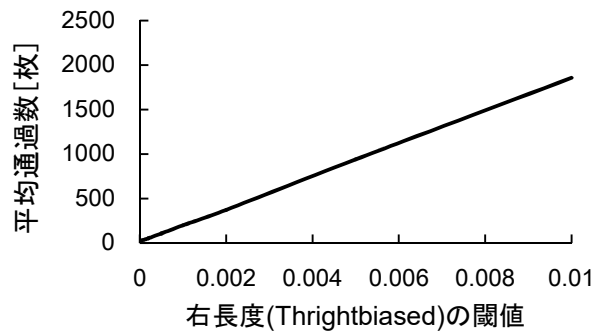
既存の手指形状推定手法では, 画像特徴量を用いた類似度計算回数は, 毎回約 300 セット程度に絞ることで, 150fps と高速な推定を可能にしていた. 本システムにおいても, 類似度計算の回数を 200～300 セット程度にすることで, 従来システムと同等の速度を実現することを目標とする. 図 3.8 より, 各閾値を個々に変化させた場合, 第 1 次探索を通過する平均の枚数はリニアに減っている. しかし, 1 つの閾値を下げるだけでは, 閾値を非常に小さくしなければ, 平均通過枚数を十分に減らすことは難しい. そこで, 図 3.9(a)のように 3 つの閾値を適切に変えることで, 1 つの閾値を極端に小さくすることなく, スクリーニングを通過する平均データセット数を減らすことができる. ただし, 閾値を小さくし過ぎると, スクリーニングを 1 セットも通過せず, 推定に失敗するという問題が発生する. 図 3.9(b)は推定毎に 1 セットもスクリーニングが通過しない確率を示したものである. 閾値を 0.01 よりも小さくした時に, 急激に確率が上がっていることが分かる. そこで, 今回の評価実験では, 平均通過枚数が従来システム程度に減らすことができ, かつ, スクリーニングを毎回 1 枚以上通過するように, 各閾値を 0.01 として評価実験を行った.



(a)縦長さの閾値(Th_{tall})を変化させた時

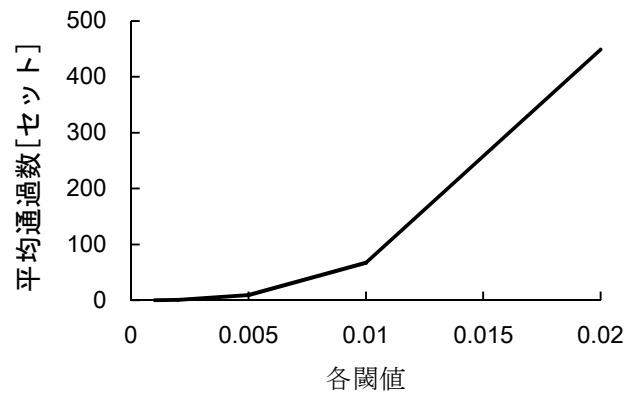


(b)上長さの閾値($Th_{topheavy}$)を変化させた時

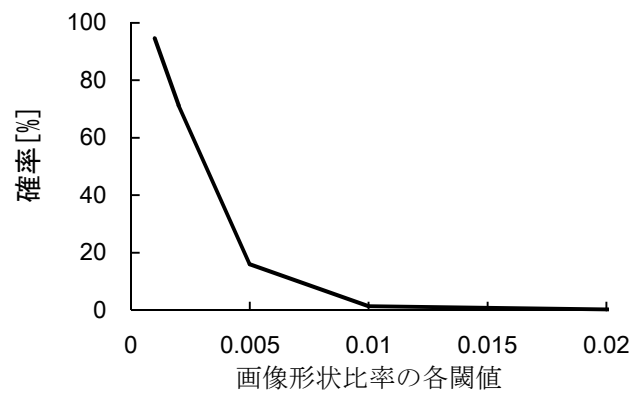


(c)右長さの閾値($Th_{rightbiased}$)を変化させた時

図 3.8 各閾値を変えた時の画像形状比率によるスクリーニングの平均通過数



(a) 閾値とスクリーニングを通過する平均データセット数



(b) スクリーニングに失敗する確率

図 3.9 3つの閾値を同時に変化させた時の画像形状比率によるスクリーニングを通過する平均データセット数と推定失敗確率の関係

3.3.2 評価実験

評価実験では、図 3.10 のようにカメラから 50cm 離れた位置を中心に 30cm 程度の範囲内で手指を動かしながら推定実験を行った。使用したカメラは Point Grey 社製 Firefly MV (FOV 60 度、解像度 640x480、撮像速度 60fps) と、PC は CPU : Intel(R) Core(TM)2 Duo CPU E6850 3.00GHz、メモリ : 2.00 GB のスペックのマシンである。使用したデータベースは、一般成人男性の手で作成し、前腕回旋角度がカメラに対して掌が向いた状態から手の甲が向いた状態までの 180 度分に関して、28,386 セットのデータセットを保持している。また、画像形状比率によるスクリーニングの閾値は、3.3.1 の予備実験で決定した 0.01 としている。

今回の評価実験では、2 種類の方法で推定精度の検証を行った。第 1 の評価実験は、推定結果を CG で描画した手で出力し、その時の手指画像と比較を行うことで、推定精度の検証を行った。第 2 の評価実験では、推定された関節角度とデータグローブで測定した。その時の関節角度の比較を行うことによって推定精度の検証を行った。この実験では、被験者にデータグローブを装着し、さらにその上から白い手袋をつけてもらった状態で、カメラの前で手指を自由に動作させてもらい、その時の推定した関節角度とデータグローブで測定した関節角度を取得した。また、この評価実験は、データベース作成を行った人物とデータベース作成を行った人物とは別の人物に被験者となってもらった。

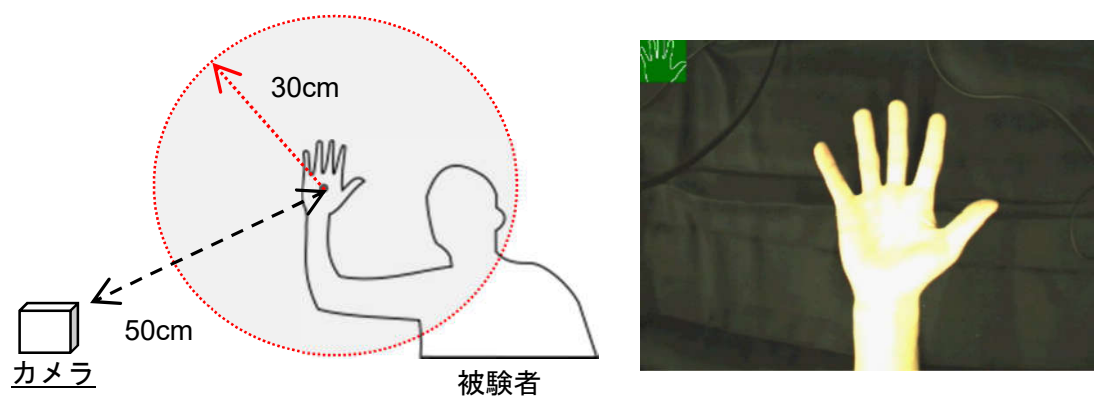


図 3.10 評価実験の撮像条件（左）とカメラから取得できる映像（右）

第 1 の評価実験を行った際のスナップショットが図 3.11 である。様々な手指形状に対して良好な推定ができていることが分かる。

第 2 の評価実験の結果を図 3.12, 図 3.13 に示す。図 3.12 の被験者は、データベース作成を行った人物, 図 3.12 はデータベース作成を行った人物とは別の人物の手による実験結果である。また, 図 3.12, 図 3.13 の(a), (b)は掌をカメラに向けた時, (c), (d)は小指側の手側面部分をカメラに向けた時で, (a), (c)は示指 PIP 関節, (b), (d)は拇指 MP 関節の実験結果である。前腕を自由に回旋させ, 指を自由に動作させた時の全指の平均推定誤差は, 誤差の平均値 -2.11 度, 標準偏差 ± 14.14 度となった。また, 推定の際, 第 1 探索の通過枚数は全データセット 28,386 枚のうちの平均 137.7 枚となり, 処理速度は平均 80fps 程度となった。本手法での推定速度は, Hoshino, et al. [20]–[26]の手法の 160fps よりも遅いように思われるが, 従来手法は手指位置をカメラに対して固定する必要があったのに対して, 本手法は手指位置を検出するアルゴリズムを追加している。さらに, 従来手法が掌がカメラに向いている状態だけであるのに対して, 本手法では 180 度の前腕回旋を許容したシステムとなっている。そのため, 単純に比較することはできないが, 本手法は従来手法と同等の速度を実現していると考えられる。また, 推定精度は, 掌をカメラに向けた状態のみであれば, 誤差の平均値と標準偏差は -2.27 ± 9.5 度であり, 従来手法の 2.45 ± 10.91 度よりも推定精度が向上している。これは, 手指領域を切り出すことによって, データベースの保持する手指形状とのマッチング精度が向上したためであると考えられる。

図 3.12(a), (b)より, カメラに向いている時の推定精度は非常に良いことが分かる。しかし, 図 3.12(c), (d)より, カメラに対して掌が横向きの時の推定精度は図 3.12(a), (b)に比べて悪いことが分かる。これは, 掌が横を向いた時には自己遮蔽が大きくなるからであると考えられる。示指から小指までの 4 指は 1 列に並んでいるためにどの指が曲がっているかの判別が難しい。また, 拇指も屈伸の状態に関わらず, 見えが変化しないためにその判別は難しいため, 推定精度が悪くなったと考えられる。さらに, データベース作成者とは違う被験者による推定結果の一例を図 3.13 に示す。本手法では, サンプルとなる人物の手を用いてデータベースの作成を行うために, データベース作成者以外の人物の手で推定を行った場合, データベース作成者に比べて推定結果が悪くなってしまうことが考えられる。しかし, 実際には推定精度は落ちるものの, 指の太さや長さ, 掌の形状の違う被験者の手を用いた時にも正確に推定を行っていることが分かる。

ただし, 現状では, データベースを作成した手指と全く違う指の太さや長さの手指を入力画像とした場合, 推定精度が落ちるという問題がある。理論上, 本システムはあらゆる指の太さや長さの手指に関する手指形状のデータをデータベースに登録することで, あらゆるユーザの手指形状を推定することが可能である。しかし, 実際には, 手指形状の個人差は非常に大きく, 例えば, 男女の手指では太さや長さが大きく異な

り，完全に不特定多数のユーザに対応するためには，現在のデータベースの何十倍もの規模のデータが必要になる．そのため，現段階では，様々な手指の太さや長さ不特定多数のユーザ全てに対して，高精度の推定を実現することは難しい．

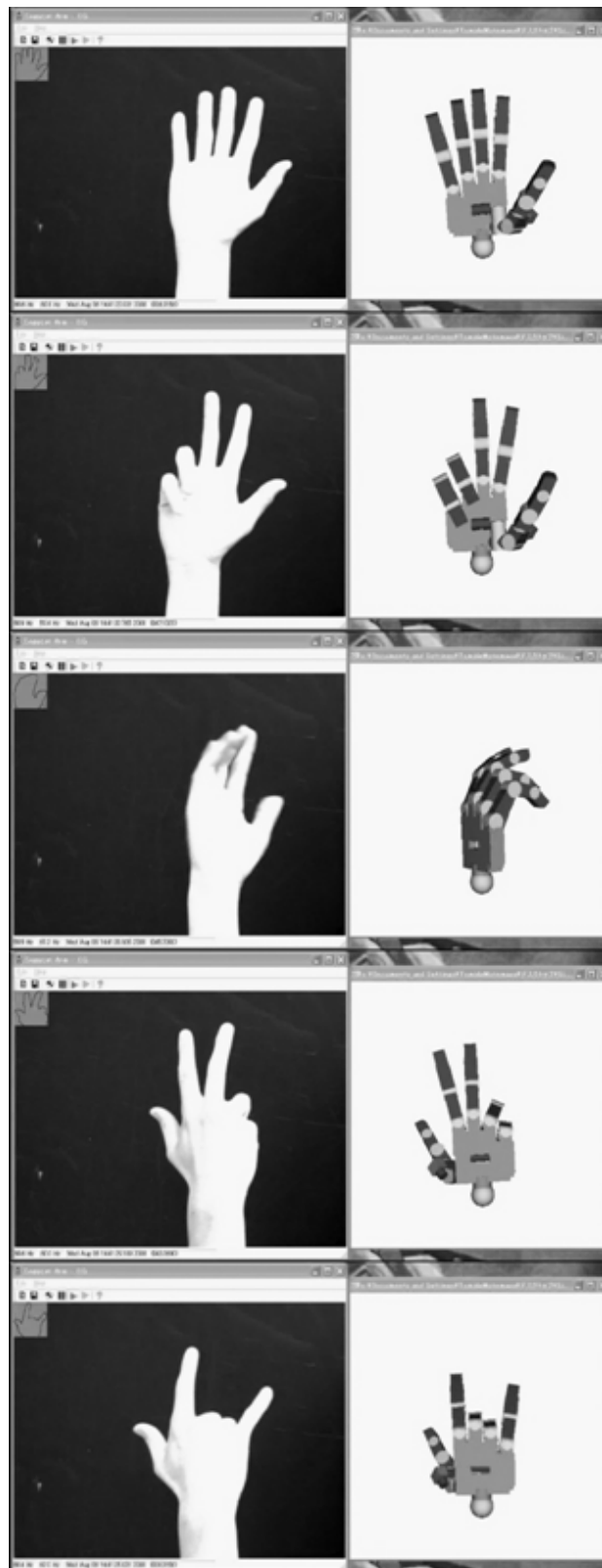
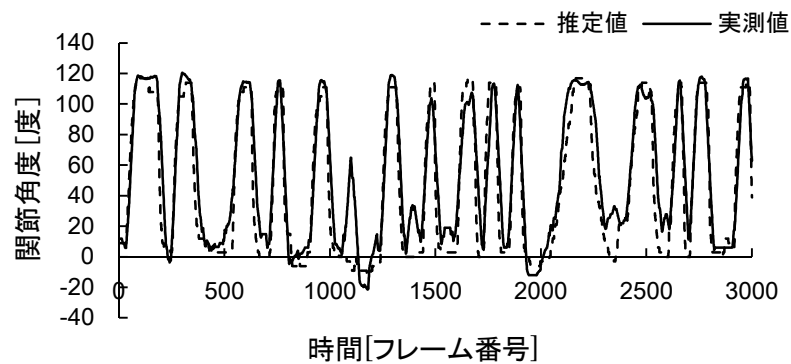
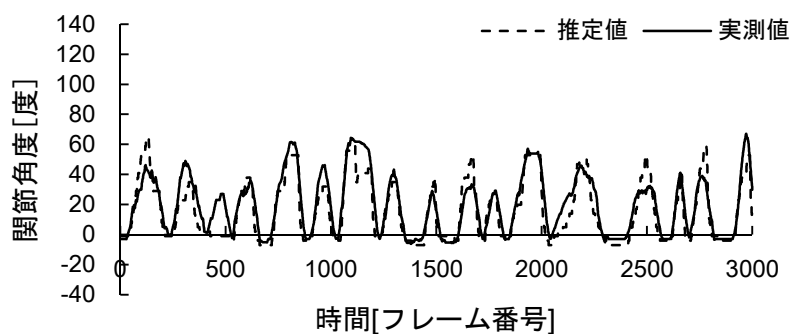


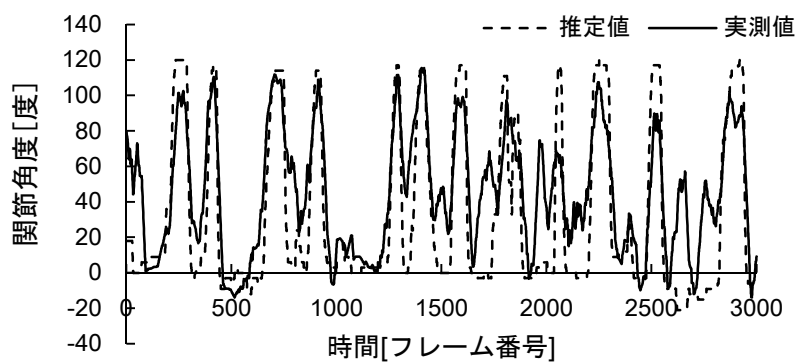
図 3.1 1 CG による手指形状推定の描画.



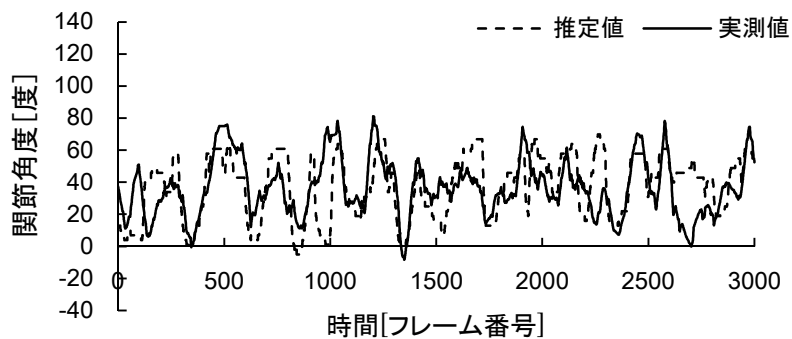
(a) 回旋角度 0 度，示指 PIP 関節推定結果



(b) 回旋角度 0 度，拇指 MP 関節推定結果

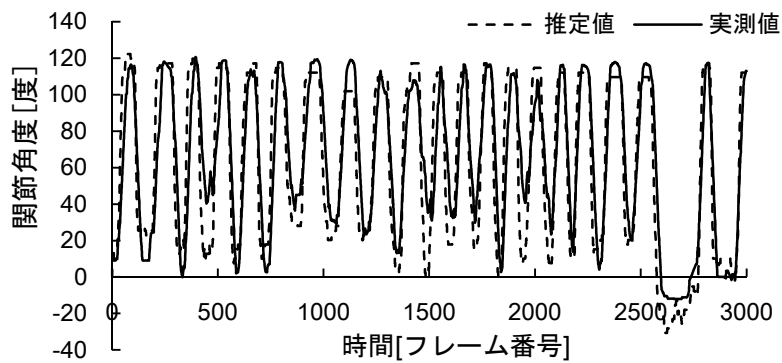


(c) 回旋角度 90 度，示指 PIP 関節推定結果

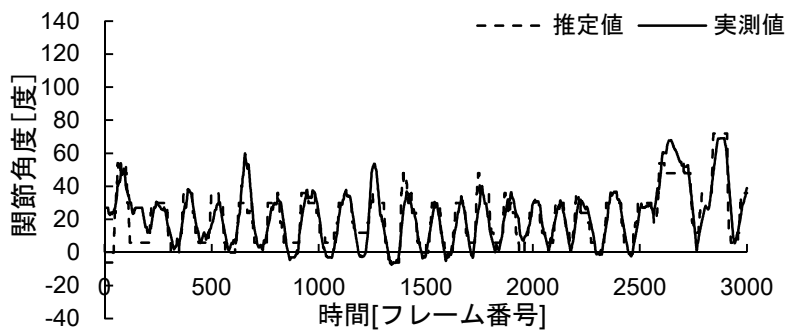


(d) 回旋角度 90 度，拇指 MP 関節推定結果

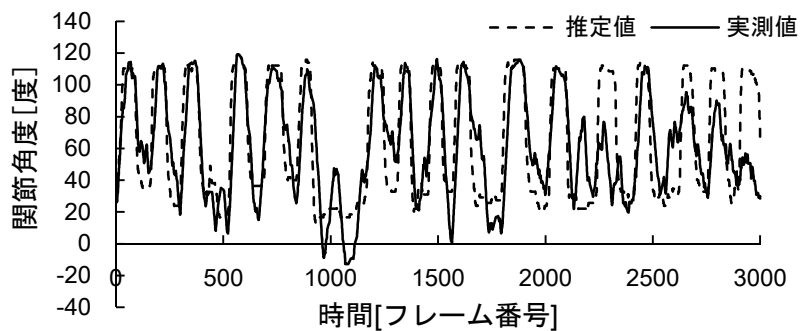
図 3.1 2 データベース作成者の手による形状推定結果例



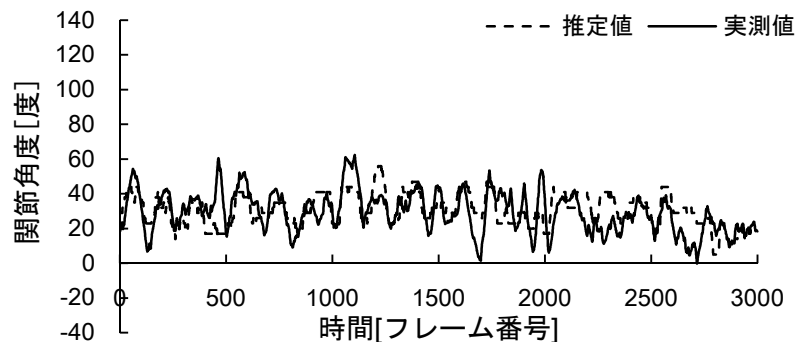
(a) 回旋角度 0 度，示指 PIP 関節推定結果



(b) 回旋角度 0 度，拇指 MP 関節推定結果



(c) 回旋角度 90 度，示指 PIP 関節推定結果



(d) 回旋角度 90 度，拇指 MP 関節推定結果

図 3.13 別の被験者の手による形状推定結果例

3.4 まとめ

従来の手指形状推定手法では，データベースの探索範囲を狭める必要があったが，特徴量空間での類似度による並べ替えを，実際の手指形状の類似度に最適化することが難しく，適切にデータベースの探索範囲を狭めることは困難であった．仮に，先験的知識を用いてデータベースを多階層化しようとした場合でも，代表的な値を持つ各階層の各クラスに，画像特徴に関するどのような情報を割り当てたら良いかを決めるのが容易でなかった．そのため，手指動作によっては，本来推定されるべき真値となるデータが探索範囲外に存在するという問題が発生してしまう．その場合，真値となるデータを出力することはできず，データベース探索範囲内の最も類似するデータを出力してしまうため，推定誤差が大きくなってしまう．さらに，高速化のためにデータベースの探索範囲が固定長であるので，一度でも真値となるデータが探索範囲外に存在してしまった場合，基本的には，それ以降の推定でも真値が探索範囲外に存在することになり，推定誤差の大きい状態が続くという問題があった．

そこで，本手法では，手指画像から低次の粗い画像情報と高次の精細な画像情報算出し，第1段階を低次の情報によるスクリーニング，第2段階を高次の情報による類似度計算とする2段階のデータベース探索を採用することで，データベースの探索範囲を狭める必要のない手法を実現する．この手法では，低次の粗い画像情報として，大まかな前腕回旋，4指の屈曲伸展と拇指の屈曲伸展を反映する3つの比率情報によってデータベースを高速にスクリーニングすることで，データベースの探索範囲を狭めることなく，高速な手指形状推定を実現する．また，この手法であれば，従来のような適切なデータが探索範囲外にある問題や過去の推定結果の影響を受けるといった問題を解決することができる．

評価実験の結果，手首関節の屈伸や回旋を含む3次元動作の手指形状推定を，誤差平均値 -2.11 度，標準偏差 ± 14.14 度，処理速度 80fps 程度と高速高精度に実現することができた．また，推定精度が高くなる掌がカメラに向いている状態であれば，誤差の平均値 -2.27 標準偏差 ± 9.5 度であり，従来手法の 2.45 ± 10.91 度よりも推定精度が向上することができた．

4 章 ゆるい直交関係にある 2 台のカメラを使った手指形状推定

システム

4.1 はじめに

ロボットによる把持動作を実現するためには、4 指と拇指の対向動作を高精度に推定できることが欠かせない。しかし、把持動作は、他の手指形状よりも自己遮蔽の影響が大きく推定が非常に難しい。特に、1 台のカメラを用いた手指形状推定システム[3]–[6][9]–[25]では、図 4.1 のように、同じ手指のシルエット画像であっても手指形状が違うために、4 指と拇指の対向動作を高精度に推定することは難しい。また、仮に、最も推定精度の良い手の甲側から撮像した場合、特に重要となる拇指動作の自己遮蔽が非常に大きくなり、肝心な部分の精度が下がってしまうという問題がある。しかし、把持動作には握力把持と精密把持の少なくとも 2 種類の姿勢が存在するため、これらの動作を正確に推定できる必要がある。また、指と把持対象との位置関係が不正確なら、対象の人工物は容易にロボットハンドから抜け落ちてしまうと考えられる。

しかし、手指形状推定技術の応用例を考えれば、マルチカメラシステム[7]のように操作対象を囲むようにして多数のカメラで撮像するのは現実的でない。また、可能なら、カメラの設置位置も厳密に指定することなく、ゆるい直交関係にあるような位置に 2 台ほどのカメラを設置する程度で済ませたい。

そこで、本章では、直交関係にある 2 台のカメラを使い、撮像されている手指面積の大きい方のカメラから 4 指の手指形状、小さい方のカメラで拇指対向の手指形状を推定するシステムを提案する[28][29]。このシステムでは、大まかな直交関係を持って設置された 2 台のカメラにより、通常スペックのパソコン 1 台の使用で、遠隔ロボットを実時間操作できる程度で高速で高精度の 3 次元手指形状推定の実現を目的とする。ゆるい直交関係にある 2 台のカメラを使うことによって、1 台では死角となってしまう部分についても高精度に推定することができる。また、通常、複数のカメラを使う場合には、カメラ同士の厳密な相対位置を検出しておく必要があるが、本手法では厳密な相対位置の検出の必要のない推定手法を使うことで、簡単にシステムを設置することができるシステム構築を目指す。

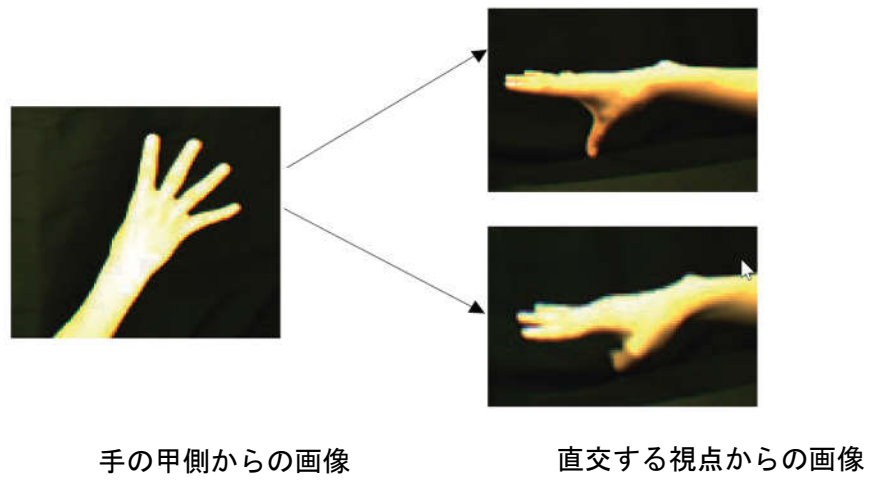


図 4.1 自己遮蔽により拇指形状の違いが分からない例

4.2 システム構成

4.2.1 2 視点からの画像情報を持つデータベース構造

本システムは, 図 4.2 のように, 直交する 2 視点にカメラを設置することを想定し, 図 4.3 のような直交する 2 視点からの画像情報を持つデータベースを持つ. 各データセットは, それぞれの視点からの画像特徴量と画像形状比率, 手指形状を表す角度情報(各指関節角度, 手首関節の屈伸と内外転角度, 前腕回旋角度)で構成されている. 画像形状比率と画像特徴量は, 各視点からの画像に 3.2.2~3.2.4 の処理を施して算出される.

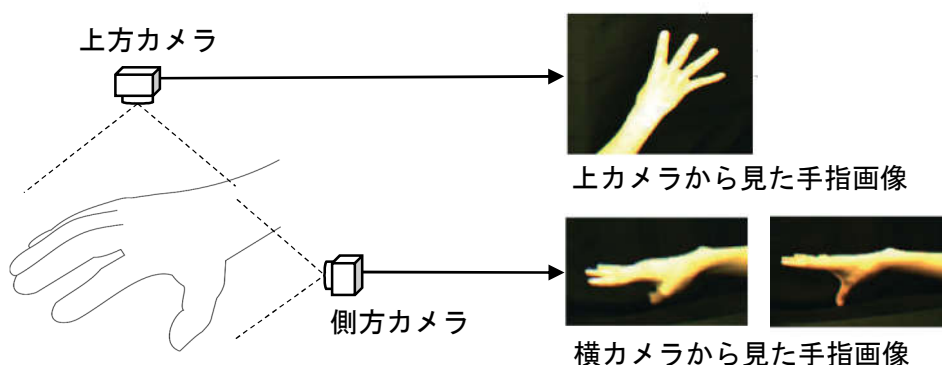


図 4.2 2 視点からのカメラ構成概要

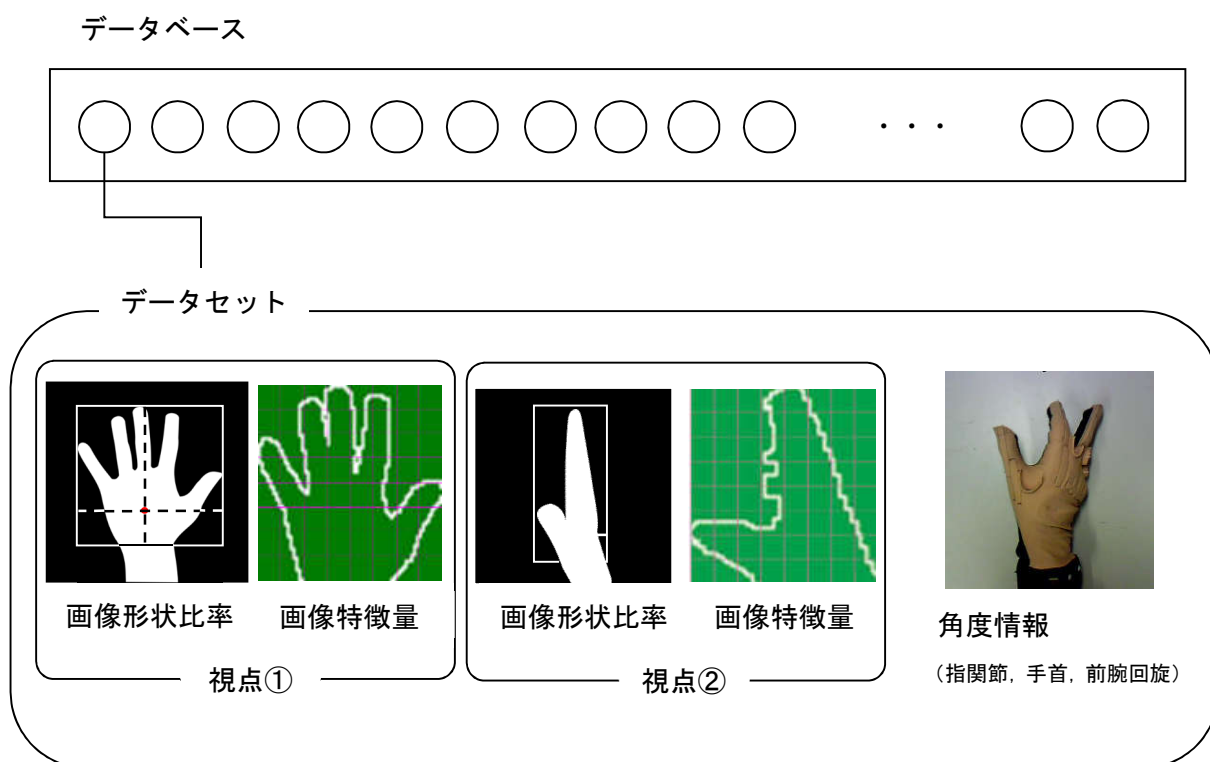


図 4.3 2 視点からの情報を持つデータベース構造

4.2.2 推定手法

ゆるい直交関係を持って設置された2台のカメラによる手指形状推定方法の基本的な発想を図4.4に例示する. 第1に, 2台のカメラで撮像された手指領域のそれぞれの面積を比較し, 面積の大きい側を決定する. 選ばれた一方の画像のみの手の形状比率を用いて, 粗いスクリーニングを行う. 簡単に言えば, 第1の処理により, 手の甲側から見た場合の大雑把な姿勢が決定される. そして第2に, 選ばれた候補のなかで, ゆるい直交関係にあるカメラにより得られた画像特徴量のみを用いて, 精緻な類似度照合を行う. 簡単に言ってしまえば, 第2の処理により, 横から見た画像を使って, どのくらい指が曲がっているかが確定される.

実際の処理としては, まず, 前処理として, 上方カメラと側方カメラの2視点から取得した画像に3.2.2~3.2.4の処理を施して, それぞれの視点からの画像特徴量と画像形状比率を算出しておく. 次に, 第1段階のデータベース探索として, 入力画像と各データセットの面積の大きい側の画像形状比率と画像特徴量を使って, 3.2.6に記述した2段階のデータベース探索手法で類似度を算出する. この際, 類似度がある閾値以下のデータセットを推定結果の候補として抽出する. 最後に, 抽出されたデータセットに対して, 面積の小さい側の画像形状比率と画像特徴量を使って, 3.2.6のデータベース探索手法を施し, 最も類似するデータセットを推定結果として出力する.

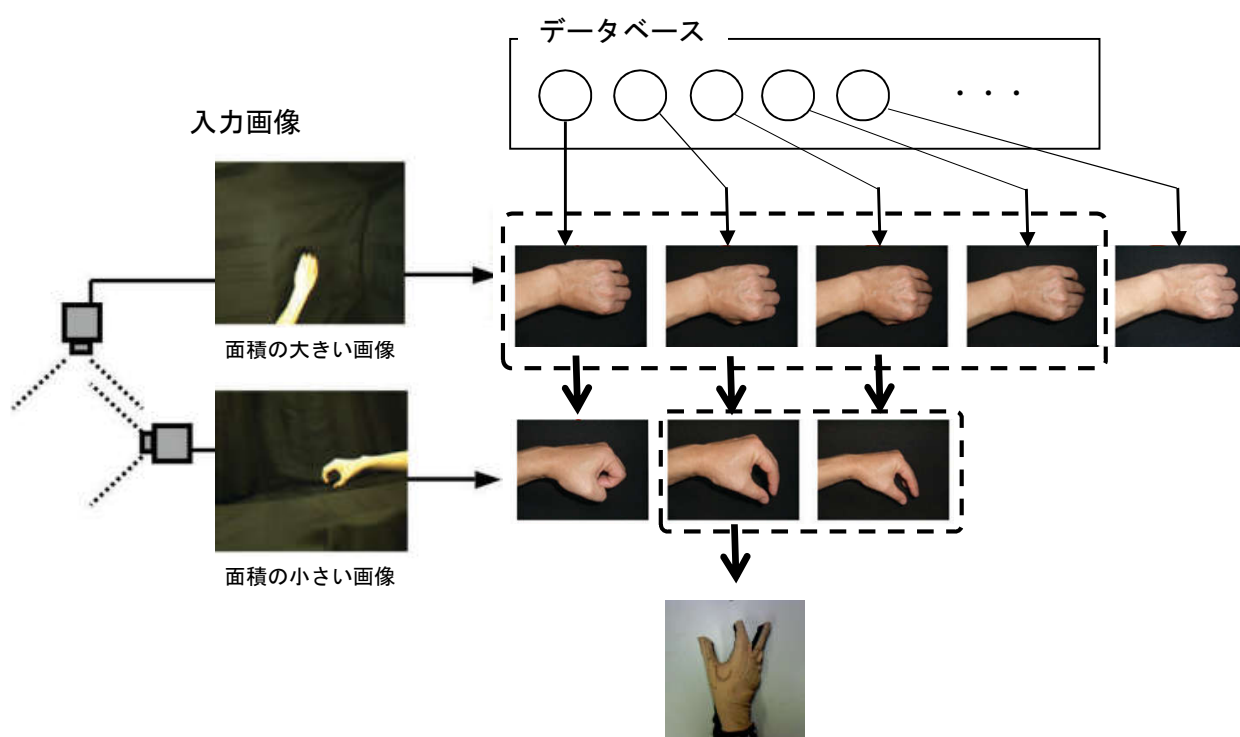


図4.4 2視点からの画像情報を使ったデータベース探索

4.3 評価

評価実験では、図 4.5 のようなカメラ配置のシステムを使った。また、使用した機材はカメラは Point Grey Research 社製 Dragonfly ExpressTM を 2 台使用し、CPU が CoreTM 2 Duo Processor T8300 (2.40 GHz, 800 MHz FSB), メモリ 4GB のスペックのノート PC を使用した。また、データベース作成のための入力画像作成には、3DCG モデリングソフトの Poser5 を使用しており、腕回旋を 7 段階とする約 30,000 のデータセットをデータベースに登録している。

また、推定精度の検証方法として、入力画像と推定結果の CG の比較とデータグローブ (Virtual Techonogies 社製 Cyber Glove) による実測値との比較の 2 種類を実施した。データグローブを使った比較実験では、右手で本システムによる手指形状推定、左手でデータグローブでの実測を同時に行い、左右の手を注意深く同期させて動作させることで推定精度を検証した。

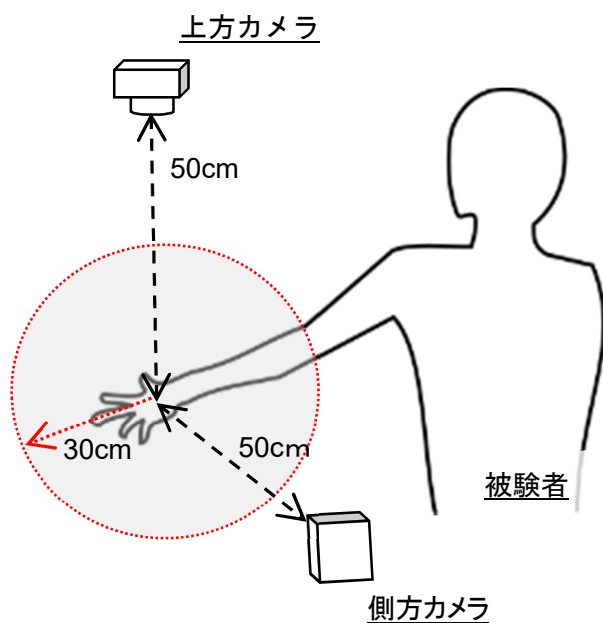


図 4.5 評価実験環境の概要

まず、図 4.6 で入力画像と推定結果の CG を並べたスナップショットにて推定結果を示す。図 4.6 では各視点からの入力画像と、その視点から見た推定結果の CG を並べて比較し、直感的に精度が分かるようになっている。この推定結果では、図 4.6 上段の2つの推定結果と下2つの結果は、上の視点からの見え方はほとんど同じ状態であるが、実際の手指形状は大きく違う姿勢を入力としている。

次に、図 4.7 では、データグローブで取得した実測値と本システムで取得した推定値を比較したグラフで推定結果を表しており、推定精度の誤差と標準偏差は、示指 PIP 関節で 0.45 ± 14.57 度、拇指 CM 関節で 4.7 ± 10.82 度であった。図 4.7(a1)と(a2)は示指 PIP 関節の推定結果、(b1)と(b2)は拇指 CM 関節の推定結果であり、(a1)と(b1)は実測値と推定値を並べたグラフ、(a2)と(b2)は推定誤差のグラフである。

直感的にも分かるように、把持動作に必要な 4 指と拇指の対向動作を含め、正確に手指形状推定できていることが分かる。例えば、図 4.6 の上の 2 つの入力画像は上方カメラから取得したが画像では同じシルエットとなってしまう2種類の手指姿勢であっても、側方カメラを使うことによって正確に拇指動作を推定できている。また、別の姿勢である図 4.6 下段の図に関しても、同様に、姿勢の違う 2 種類の把持動作を推定できている。

このシステムを使うことによって、図 4.8 のように、実際のロボットを遠隔操作し、物体把持するデモンストレーションできるシステムを実現している。図 4.8 では、本システムで推定した手指形状を推定した結果を使って人型のハンドとアームを操作している様子であり、実際にロボットが物体把持できている様子が分かる。このシステムは、SIGGRAPH などの展示会にも出展しており、安定した動作を実現できている。

ただし、現段階では、推定誤差の標準偏差が 10 度程度であり、指先を数 mm 以下の分解能で動作させる必要がある繊細な動作を推定することはできない。つまり、本システムを使うことで、ユーザは対象の物体のどこを、どのようにして掴むのか、又は握りこむのかを、状況に応じて選択し、自身の動作に合わせて操作することはできるものの、実際に掴む直前の指先が当たるか当たらないかを細やかに操作し、適切な力を加えて物体を落とさないように繊細にコントロールすることはできない。この原因は、高精度な推定を実現するために高い分解能を持つデータベースから推定するためには、計算コストが指数関数的に増大する点にある。仮に、各指を 5 パターンで表現し、前腕回旋角度を 10 段階としたとすると、 $5^4 \times 10 = 31,250$ パターンのデータベースを持っている。実際には、ヒトがデータグローブを使ってデータベースを作成し、ヒトが使わない手指形状は含まれていないため、もう少し分解能は細くなるが、各関節の分解能は 10 度程度となっている。より高分解能のデータベースを用意すると、計算コストは 10 倍、100 倍と高くなってしまう。そのため、より細かい操作を実現するためには、より大規模なデータベースから推定できるアルゴリズムが必要となる。

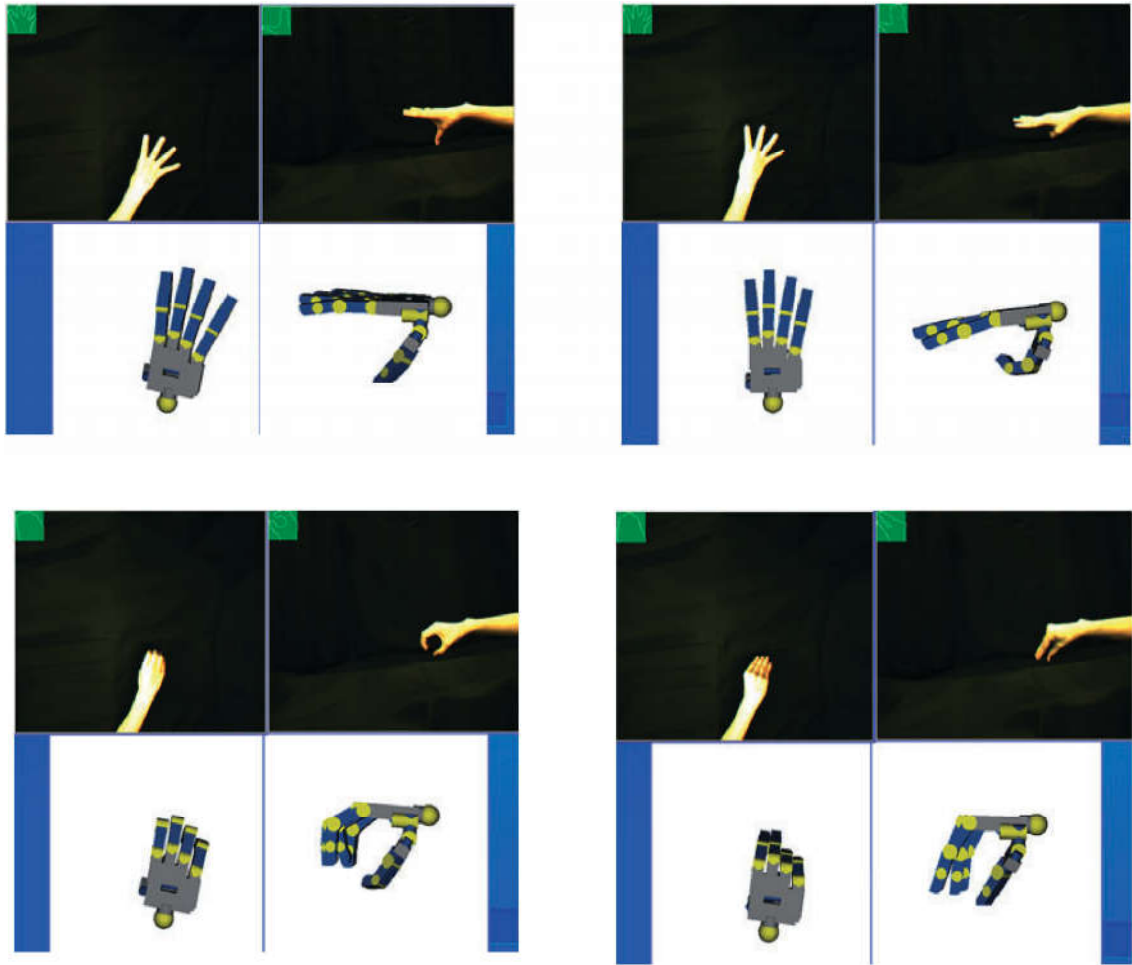
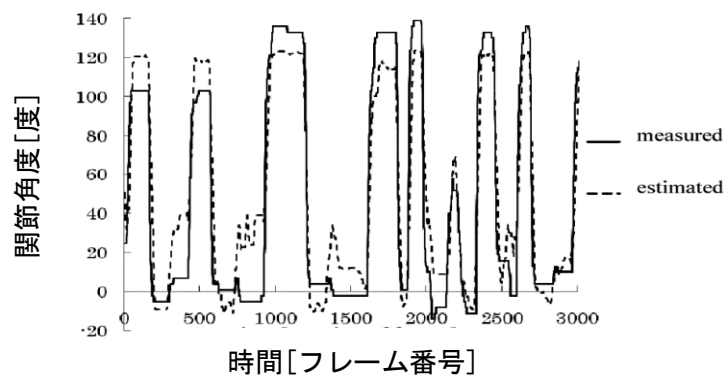
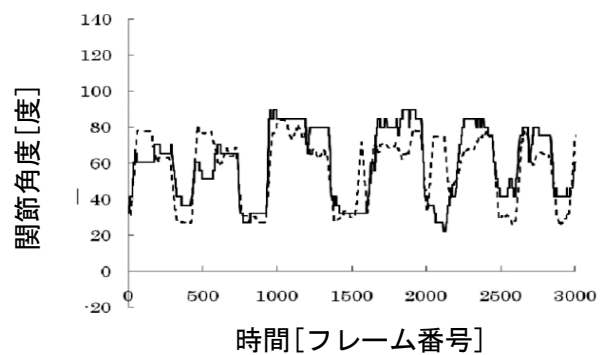


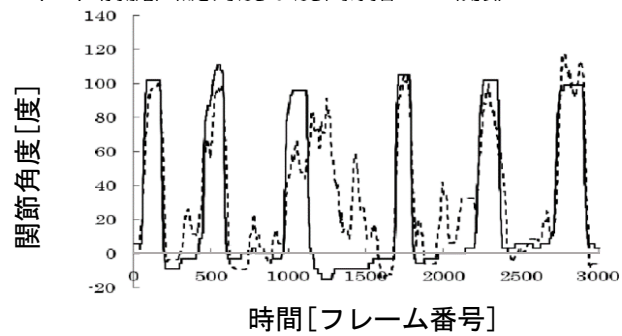
図 4.6 推定結果のスナップショット



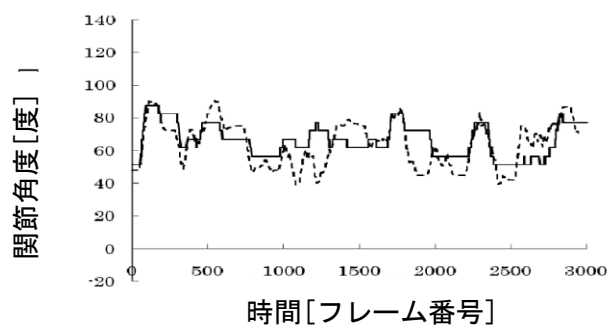
(a1) 前腕回旋角度 0 度, 示指 PIP 関節



(a2) 前腕回旋角度 0 度, 拇指 CM 関節



(b1) 前腕回旋角度 90 度, 示指 PIP 関節



(b2) 前腕回旋角度 90 度, 拇指 CM 関節

図 4.7 データグローブによる推定結果例

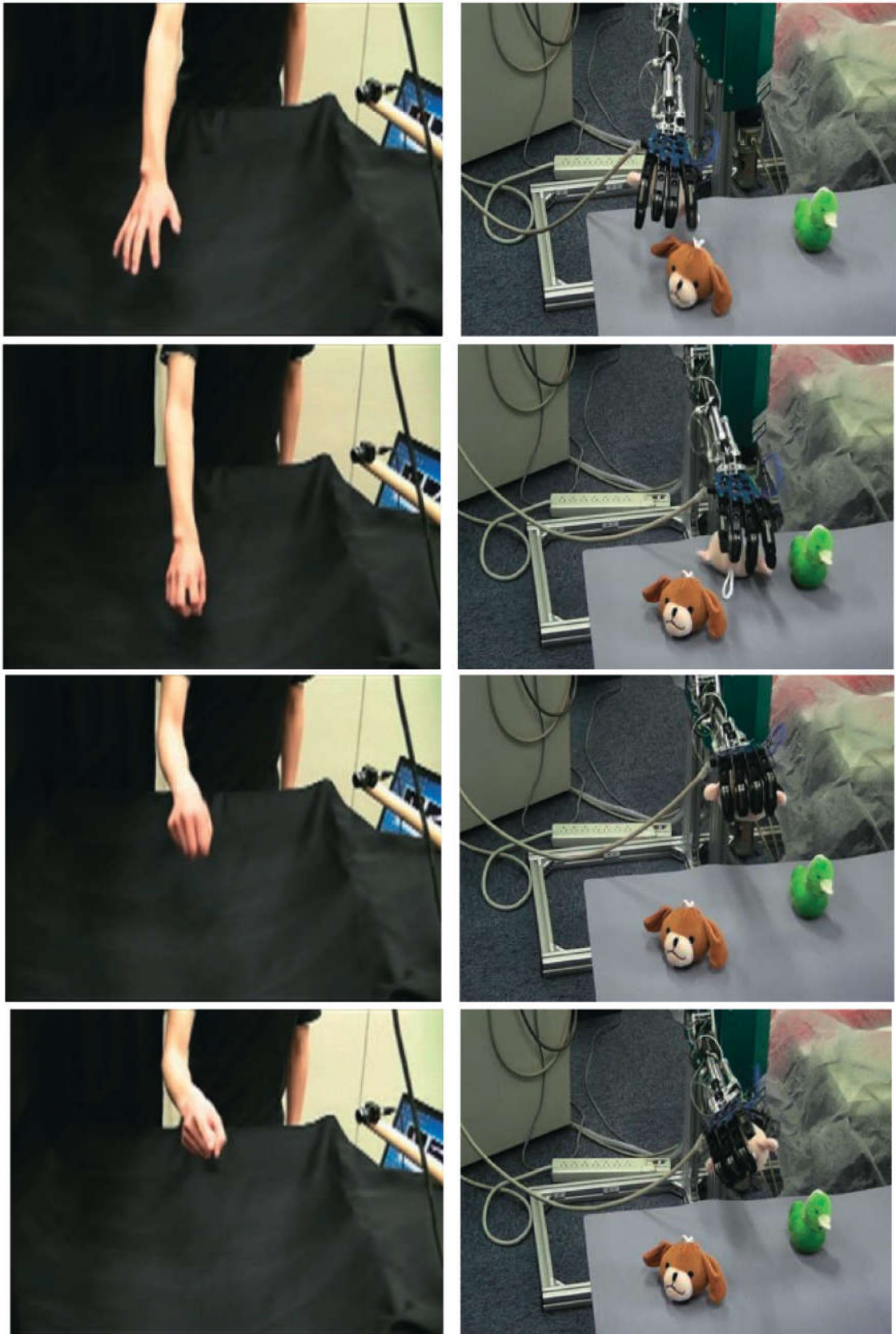


図 4.8 手指形状推定を用いた遠隔ロボット操作による安定した物体操作

4.4 まとめ

本章では、ロボットによる把持動作のために最も重要である 4 指と拇指の対向動作を高精度に推定できるシステムの実現を目指した。把持動作は、他の手指形状よりも自己遮蔽の影響の大きく推定が非常に難しい。このようなロボットシステムで把持動作を実現させるためには、自己遮蔽の影響が大きい中であっても、各指の関節角度を正確に推定できる必要がある。

しかし、1 台のカメラを用いた手指形状推定システムでは、4 指と拇指の対向動作を高精度に推定することは難しい。これは、自己遮蔽によって、同じ手指のシルエット画像であっても手指形状が違ってくるからである。しかし、マルチカメラシステムのように操作対象を囲むようにして多数のカメラで撮影するシステムでは、セットアップの手間がかかってしまい、簡単に使うことができるという有用性が失われてしまう。システムのセットアップには、カメラの設置位置も厳密に指定することなく、ゆるい直交関係にあるような位置に 2 台ほどのカメラを設置する程度で済ませたい。

そこで、本研究では、直交関係にある 2 台のカメラを使い、撮像されている手指面積の大きい方のカメラから 4 指の手指形状、小さい方のカメラで拇指対向の手指形状を推定するシステムを提案する。このシステムでは、大まかな直交関係を持って設置された 2 台のカメラにより、通常スペックのパソコン 1 台の使用で、遠隔ロボットを実時間操作できる程度的高速で高精度の 3 次元手指形状推定の実現を目的とする。ゆるい直交関係にある 2 台のカメラを使うことによって、1 台では死角となってしまう部分についても高精度に推定することができる。また、通常、複数のカメラを使う場合には、カメラ同士の厳密な相対位置を検出しておく必要があるが、本手法では厳密な相対位置の検出の必要のない推定手法を使うことで、簡単にシステムを設置することができるようにする。

評価実験の結果、示指 PIP 関節で誤差の平均値と標準偏差が 0.45 ± 14.57 度、拇指 CM 関節で 4.7 ± 10.82 度となった。また、コンピュータグラフィックスで表示した推定結果を入力画像を比較すると、従来手法では、推定が難しかった拇指をはじめとする詳細な指動作に関しても、正確に推定できていることが分かる。さらに、実機のロボットと組み合わせた試作ロボット操作システムによる実験では、ユーザがカメラの前で手指を動作させるだけで、簡単にモノを掴むことが実現できていることが確認できた。

一連の評価実験の結果から、本手法はロボットを通じた人工物操作システムの入力システムとして最適な手指形状推定システムであることが示唆される。

5 章 超小型カメラを使ったウェアラブル手指形状推定システム

5.1 はじめに

本章では、モーションキャプチャシステムのように、広い空間で自由に動き回り、様々な姿勢をとるユーザを対象とする状況に対応できる手指形状推定システムの構築を図った[30]。このシステムは、光学式モーションキャプチャシステムなどの他の入力システムと併用して使用することを想定している。例えば、光学式のモーションキャプチャであれば、簡単に頭部や腕、手といった大きい部位の動作を高精度に検出することができるものの、指のような小さく複雑な部位を、同時に高精度に検出することは難しい。そこで、指先に関しては、本研究の手指形状推定システムによって推定することで、全身の動作をキャプチャーすることがシステムができる。この時、ユーザがどんな姿勢であっても手指形状を推定できるシステムが必要となる。

しかし、従来手法[1]・[25]では、カメラ位置が空間に対して固定されているため、撮像される手指が非常に小さかったり、体で隠されてしまったりするという問題がある。また、ユーザが動いても推定可能なシステムとして、D.Kim, et al. [31]によってウェアラブルデバイスによる指動作推定システムが提案されているが、手首にデバイスを装着するために、手首を少しでも屈曲伸展させると推定することができないため、非常に限定的な用途でしか使用することができない。他のシステムと組み合わせることを想定する場合、併用する入力システム上でユーザがどのような動作をしていても、常に手指形状を推定することができるハードウェア構成が必要である。

そこで、超小型カメラと小型コンピュータによる装着型ハードウェアを使った手指形状推定システムによって、オペレータの動きを妨げずに、手指形状推定できるシステムの実現を図った。ハードウェアに使用する超小型カメラは、小さいもので 7mm×7mm 程度と非常に小型であるので、装着してもオペレータへの負担が少なく、自由な動きを妨げない。

ただし、小型コンピュータは通常の PC よりもスペックが低いため、高速に動作する手指形状推定アルゴリズムが必要となる。低スペックのコンピュータ上で、高速に推定できない原因は、データベースの持つ特徴量が 1,600 次元と高次元である点にあった。これを低次元化することができれば、低スペックの PC 上でも高速に推定することが可能となる。

また、本装置はオペレータの装置装着感をより快適にするため、小型カメラをオペレータの操作を妨げない手の甲側から、できるだけ近い位置に設置する必要があるが、手指を大きく動かすと、簡単に指先が撮像範囲外に出てしまう。そのため、画像形状比率によるスクリーニングを使用することが難しくなるため、画像特徴量の低次元化が必要不可欠となる。

そこで、本章では、特徴量を次元圧縮することで計算コストの低減させ、低スペック

な小型コンピュータ上であっても高速高精度に動作する手指形状推定システムの実現を目指した。そこで、本章では画像特徴量の算出に使用する HLAC に着目した。HLAC は、0 次相関のピクセルの存在を示す 1 パターン、1 次相関の方向付きの傾きを示す 4 パターン、2 次相関の方向づけられた曲率を示す 20 パターンで構成されているが、手指形状推定の類似度計算において 0 次相関の影響が非常に大きい。そこで、0 次相関の 1 パターンのみ使用する、つまり、手指シルエットの輪郭線のピクセル数のみを使うことで、推定精度を維持しながら、画像特徴量を 1/25 まで圧縮し、画像特徴量を 1600 次元から 64 次元に低減することを目指した。

5.2 システム構成

5.2.1 ウェアラブル入力デバイス

ハードウェアは、図 5.1 のように、超小型カメラと小型コンピュータで構成される。オペレータは、手指全体が撮像されるように、掌に超小型カメラを装着し、腰に小型コンピュータを装着する。また、本装置は、推定結果を無線通信で送信するための WiFi モジュールを搭載しており、推定結果を無線で送信する構成となっている。推定結果は、各関節角度であり、データ量が非常に小さいため、WiFi 程度の通信量でも十分に通信速度を確保することができる。

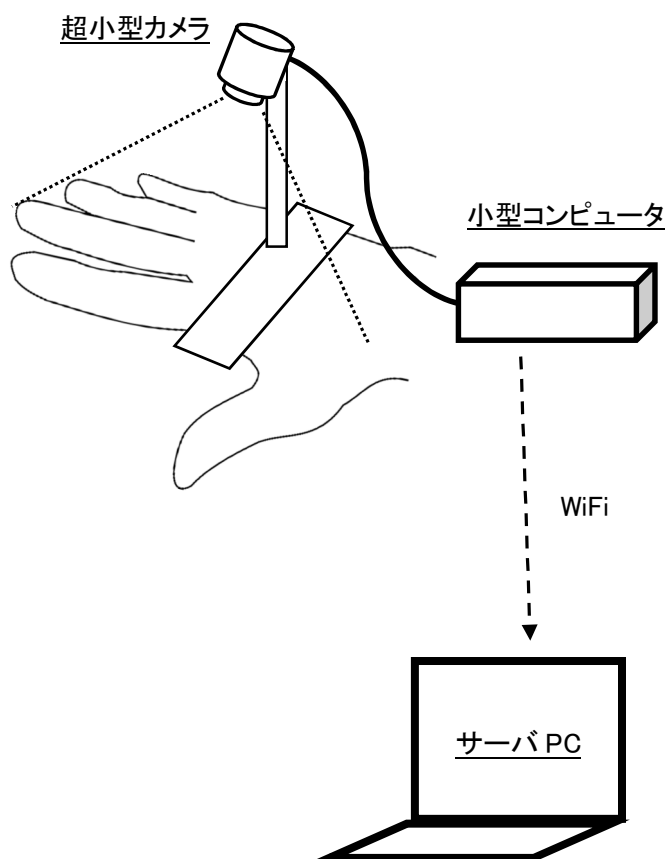


図 5.1 ウェアラブルデバイスの構成概要

5.2.2 画像特徴量の次元圧縮

画像特徴量の算出に使用している HLAC は、0 次相関(M0)のピクセルの存在を示す 1 パターンと 1 次相関(M1)の方向付きの傾きを示す 4 パターン、2 次相関(M2)の方向づけられた曲率を示す 20 パターンの計 25 パターンの 3×3 のマスクパターンで構成されている。既存の手法では、25 パターン全てを使って画像特徴量を算出していた。しかし、手指形状推定の類似度計算では、M0 の影響が他のパターンに比べて非常に大きい。

そこで、本章では、図 5.2 のように、画像特徴量の算出に、0 次相関の 1 パターンのみ使用する。具体的には、まず、2 章の手法と同様に、3.2.2 の処理で手指範囲を設定した後、3.2.4 の処理で 8×8 に分割した手指シルエットの輪郭線画像を生成する。次に、各分割画面に対して、M0 の 1 パターンのみを適用して画像特徴量を算出する。この処理によって、従来の $1/25$ の次元数である 64 次元の画像特徴量を得ることができる。

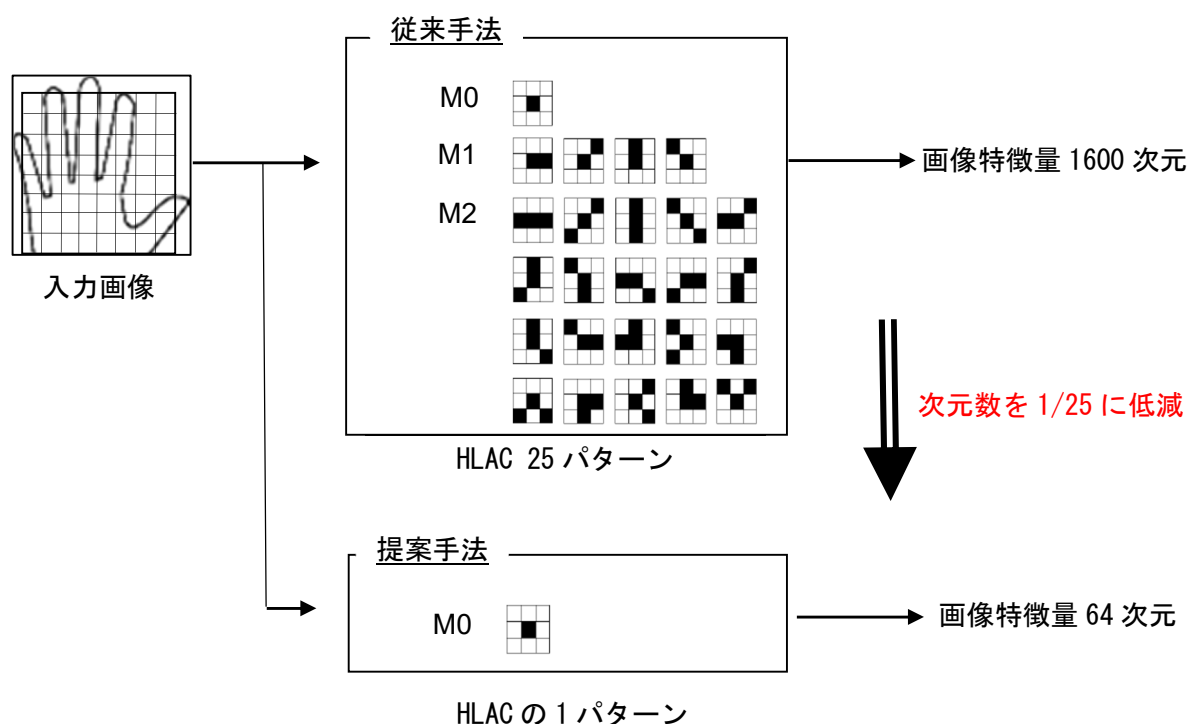


図 5.2 画像特徴量の次元圧縮概要

5.2.3 データベース構造

本章のシステムで使用するデータベースは、図 5.3 のように、圧縮した画像特徴量と手指形状を表す各種角度情報を組とするデータセットの集合で構成されている。画像特徴量は、1/25 に圧縮されているため、64 次元となっている。

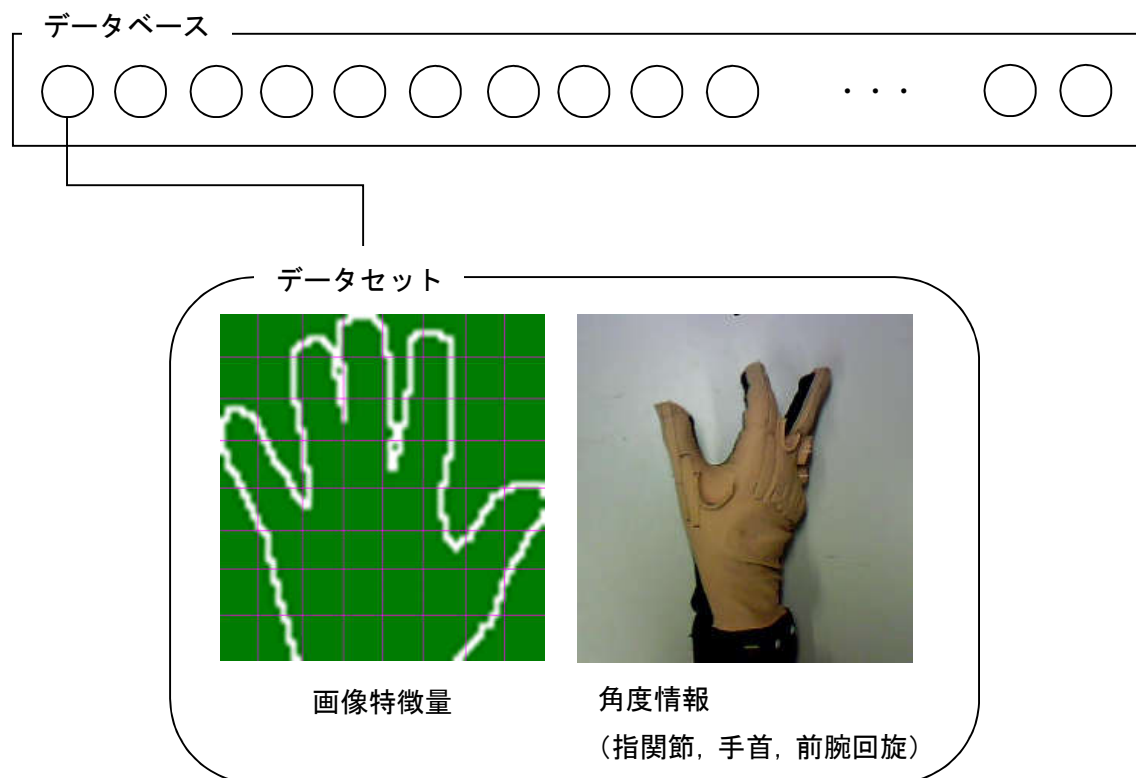


図 5.3 次元圧縮した画像特徴量を使ったデータベース構造

5.2.4 推定手法

ウェアラブルデバイスによる手指形状推定では、手指に対してカメラ位置が固定されているため、事前のキャリブレーションによって、図 3.2 で示した手指範囲を検出しておく。まず、図 5.4 上段のように、手の甲側にカメラが位置するようにデバイスを装着する。次に、カメラの方向を調整して、図 5.4 下のように手指全体が撮像されるようにカメラ方向を調整する。この時、4 指基準線上に 4 指付け根が位置し、拇指基準線上に拇指付け根が位置するようにする。最後に、手指形状を図 5.4 下のようにし、手指全体がちょうど入る範囲を手指範囲として登録することでキャリブレーションが完了する。

ウェアラブルデバイスを使った手指形状推定では、キャリブレーションで検出した手指範囲を使い画像特徴量を算出し、データベース探索を行う。具体的には、まず、設定した手指範囲内を手指領域とそれ以外で 2 値化した画像に対して、4.2.1.1 の画像特徴量の次元圧縮を適用した 2.2.4 の処理を施して、64 次元の画像特徴量を算出する。次に、4.2.1.3 で示した方法でデータベース探索を行い、再類似のデータセットが持つ角度情報を推定結果として出力する。類似度計算では、画像特徴量の次元数が 1600 次元から 64 次元に低減されているため、計算コストは 1/25 まで高速化されている。

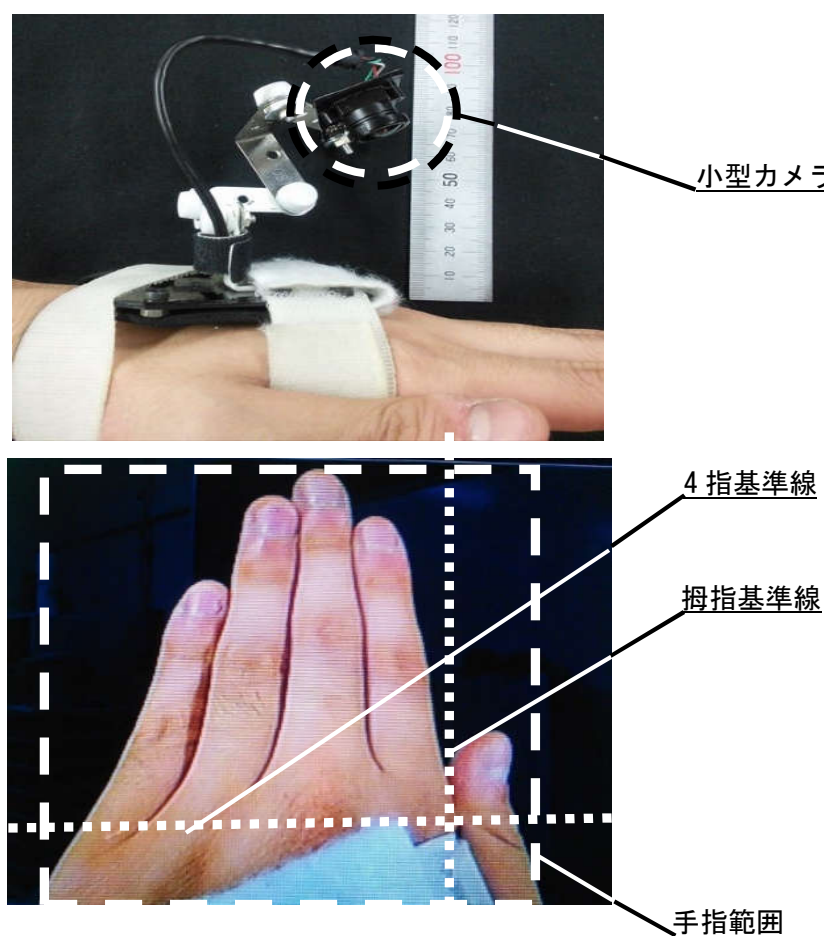


図 5.4 カメラキャリブレーション

5.3 評価

5.3.1 次元圧縮検証のための予備実験

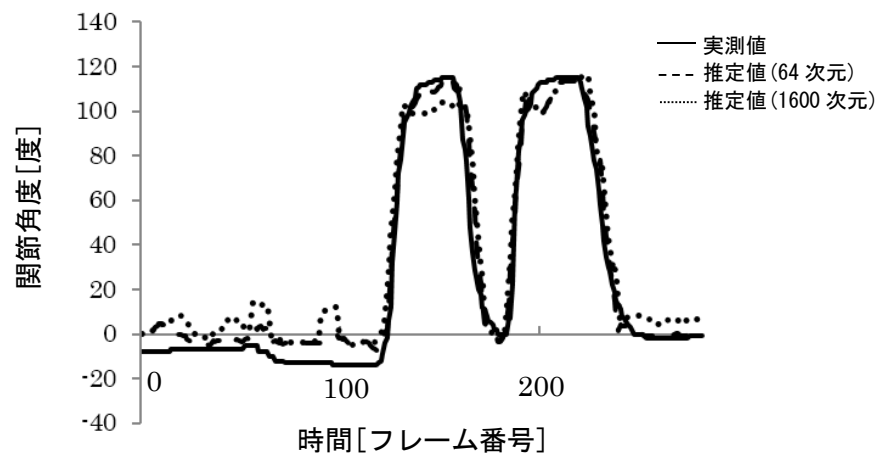
評価実験では、画像特徴量を 64 次元とするシステムの推定値と 1,600 次元とするシステムの推定値、データグローブ（Virtual Technologies 社製 Cyber Glove）の実測値を比較した。評価用の入力データは、右手にデータグローブ、左手で入力画像を取得しつつ、左右の手を同期しながら動かして作成した。また、各システムのデータベースは、評価実験と同じ方法で入力画像と角度情報の組を取得し、それぞれ画像特徴量の次元数だけ変えて作成し、全く同じ手指形状の 30,000 組のデータセットを持つ。

評価実験の結果を図 5.3 と表 5.1 で示す。図 5.3 では、実測値を太線、画像特徴量を 64 次元とするシステムによる推定値を細線、1,600 次元とするシステムによる推定値を点線で示しており、(a)は示指 PIP 関節、(b)は母指 CM 関節に対する推定結果を示している。なお、いずれも掌をカメラに向けた状態での推定結果である。同図から直感的にも分かるように、画像特徴量の次元数の違いによる推定精度はほとんどない。推定誤差の平均値と標準偏差を表 1 に示しているが、示指 PIP 関節では、画像特徴量 64 次元の場合で誤差の平均値と標準偏差が -3.92 ± 10.88 度、1600 次元の場合で -1.96 ± 8.42 度であり、母指 CM 関節に関しては 64 次元の場合で -4.81 ± 7.55 度、1600 次元の場合で -2.94 ± 7.02 度と、最大でも 2 度程度しか変わらず、ほとんど同じ推定精度であることが分かる。

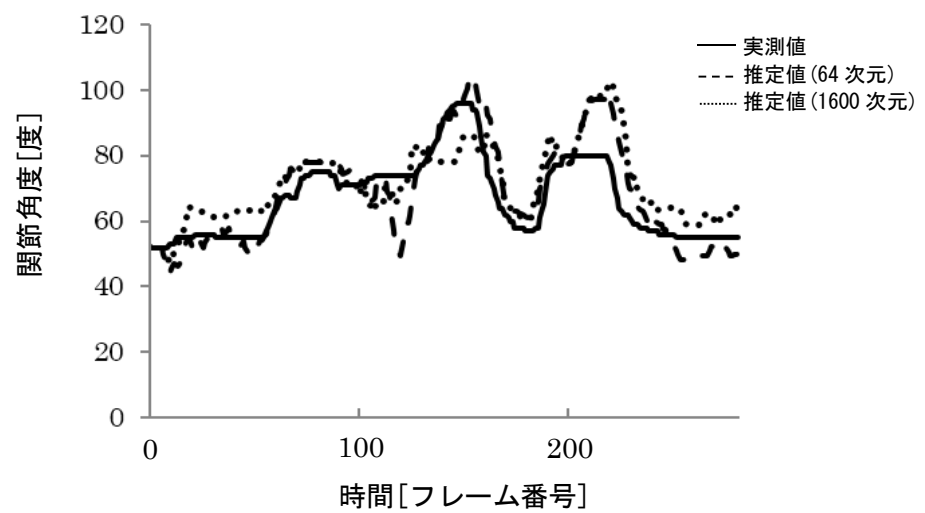
表 5.1 画像特徴量の次元数の違いによる推定誤差

次元数	示指 PIP 関節		母指 CM 関節	
	平均値	標準偏差	平均値	標準偏差
1,600	-1.96	8.42	-2.94	7.02
64	-3.92	10.88	-4.81	7.55

[度]



(a) PIP joint of index finger



(b) CM joint of thumb

図 5.5 画像特徴量の次元の違いによる推定結果

5.3.2 評価実験

評価実験では、図 5.6 に示す装置の試作機を使って、手指形状推定の精度評価を行った。試作機のカメラには、バッファローBSW20KM11BK の基板を取り出し使用しており、解像度 640x480[pixel]、撮像速度 30[fps]、画角 120 度となっている。また、小型コンピュータには、Beagle Bone Black A5C (CPU: 1GHz, メモリ: 512MB) を使用している。小型カメラはベルクロを手の甲に巻いて止める構造となっており、図 5.6 下段左のように手の甲から 7cm 程度上側に位置するようになっている。図 5.6 下段右は、カメラから取得できる画像の一例である。なお、データベースには 3,000 組のデータセットが保持されている。

実験では、左手に本装置、右手にデータグローブ (Cyber Glove Systems 社製 CyberGlove II) を装着した状態で、注意深く左右の手に同じ動作をさせて実験を行った。装置から検出した値を推定値、データグローブから検出した値を実測値として、推定精度の検証を行った。また、本評価実験でも画像特徴量 1600 次元と 64 次元での推定精度の違いを比較した。

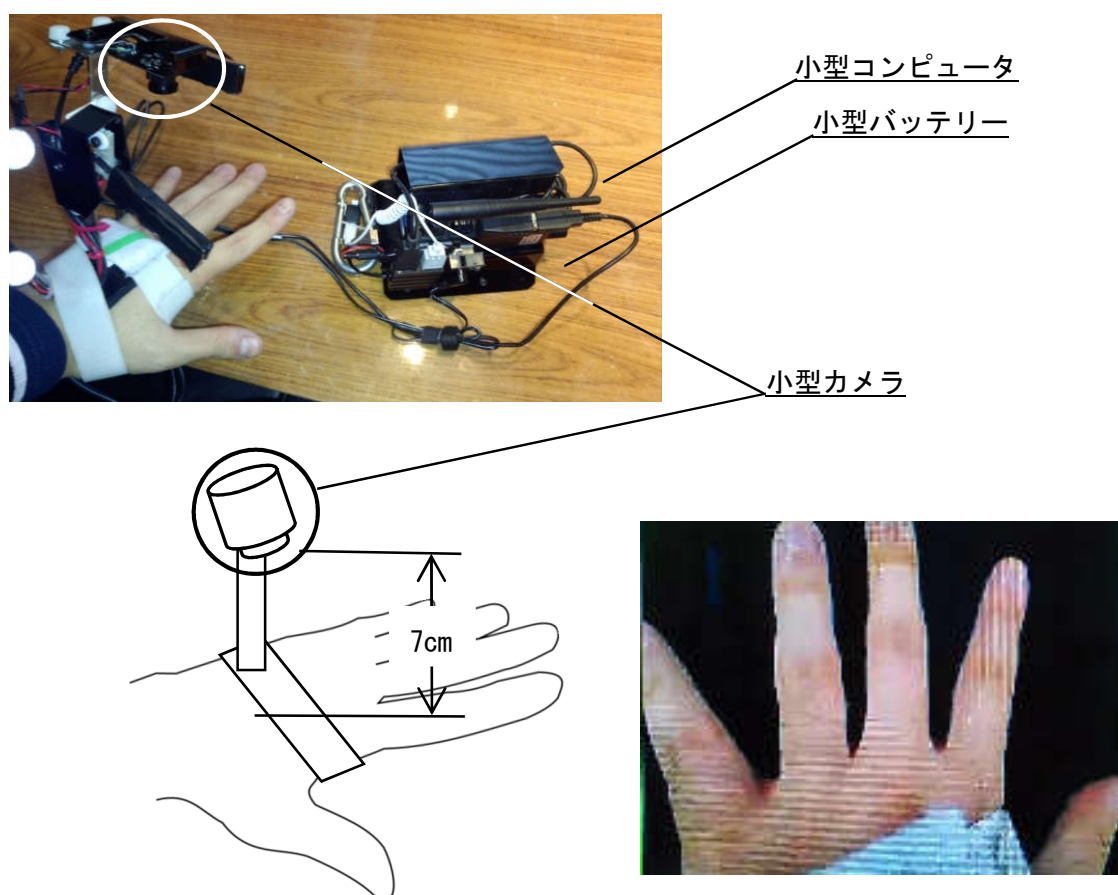


図 5.6 評価で使した試作機

評価実験の結果を、図 5.7 のスナップショットで示す。同図では、全体の推定精度が直感的に分かるように、推定結果を、右を入力画像、左を推定結果の CG とするスナップショットである。また、図 5.8 のデータグローブによる実測値との比較のグラフでは、推定誤差の平均値と標準偏差が、従来手法で -6.0 ± 6.9 度、本手法では -5.7 ± 7.2 度とほぼ同等の推定精度を実現することができた。また、推定速度は、従来手法では $158.7[\text{ms/frame}]$ に対して、本手法では $25.0[\text{ms/frame}]$ と大幅に向上させることができた。

ウェアラブルデバイスでの推定では、ユーザの自由な動作の妨げとならないように、小型カメラをできる限り手の甲に近い位置に設置するようにしている。そのため、図 5.6 でも示しているように、手指を広げると、簡単にカメラの撮像範囲外に指先が位置してしまう。しかし、図 5.8 において、0 度付近が指を伸ばした状態に対応するが、指先が撮像範囲外に位置することによる推定精度の低下はほとんど見られなかった。むしろ、指が曲がった状態である 80 度付近の方が、やや推定精度が落ちているように見える。これは、指が伸びた状態の場合、指先が撮像されなくなってしまうとしても、撮像範囲内にある指の側面などから手指形状が容易に推定できるからであると思われる。むしろ、3 章で問題となった点と同じく、握りこんだときの自己遮蔽による推定精度の低下の影響が大きいと考えられる。つまり、本手法は、推定精度を低下させることなく、ウェアラブルデバイスによる手指形状推定システムが実現していると言える。また、握りこんだ状態での推定精度を向上させるには、3 章のように直交する 2 箇所カメラを設置することで、高精度化が可能であることが示唆される。

ただし、本システムは、2 章で述べた画像形状比率を使ったスクリーニングを組み合わせることで、より大規模なデータベースを使い推定精度を向上させることは難しい。これは、小型カメラを手の甲に近づけているために、多くの手指形状で指先が撮像範囲からはみ出してしまい、適切に画像形状比率を算出することができない点にある。そのため、適切に画像形状比率を算出することができず、十分にスクリーニングすることができない。逆に、常に手指全体を撮像しようとする、小型カメラを手の甲から 10cm 以上離す必要があるが、そうするとユーザの動作の妨げとなってしまう有用性を下げる結果となってしまう。現状では、ウェアラブルデバイスでの手指形状推定手法では、画像特徴量の次元圧縮のみを使った本手法が最適であると考えられる。

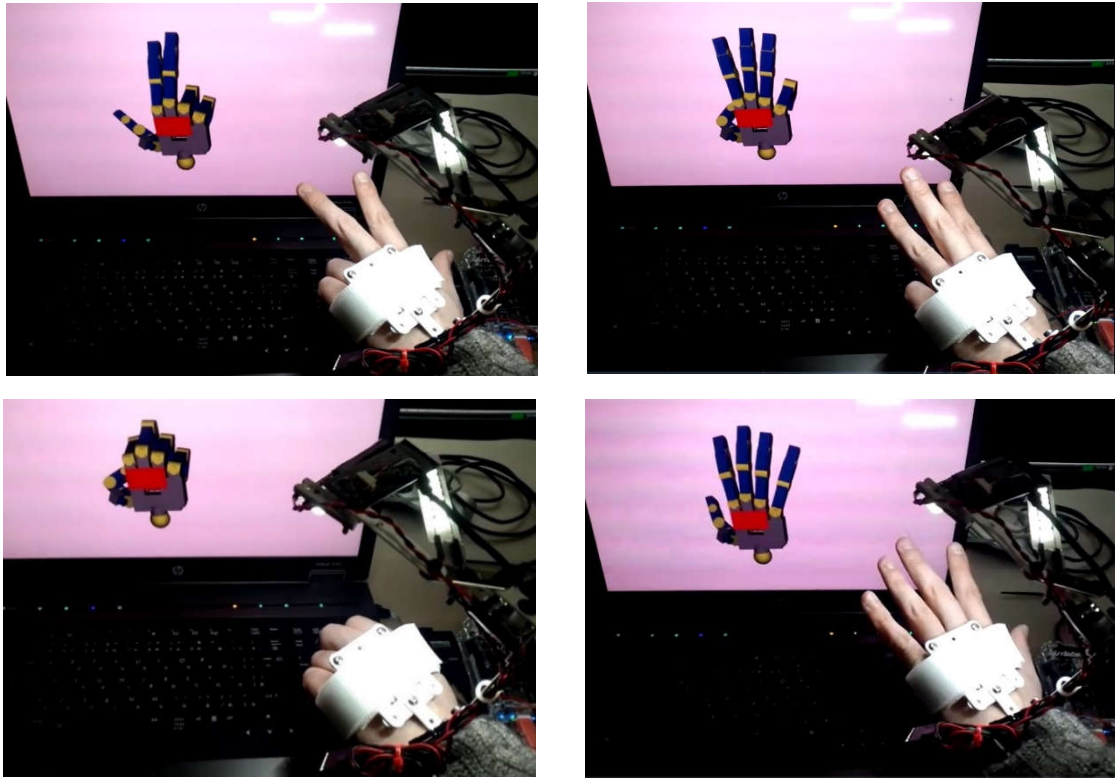
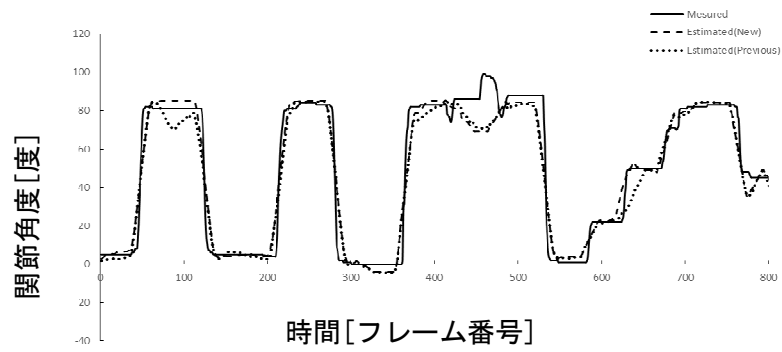
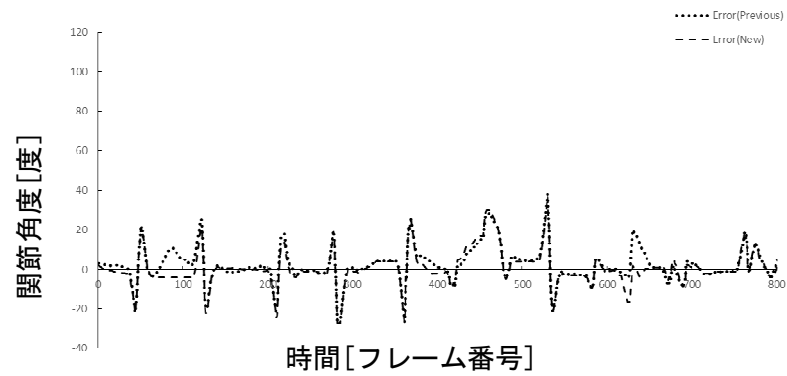


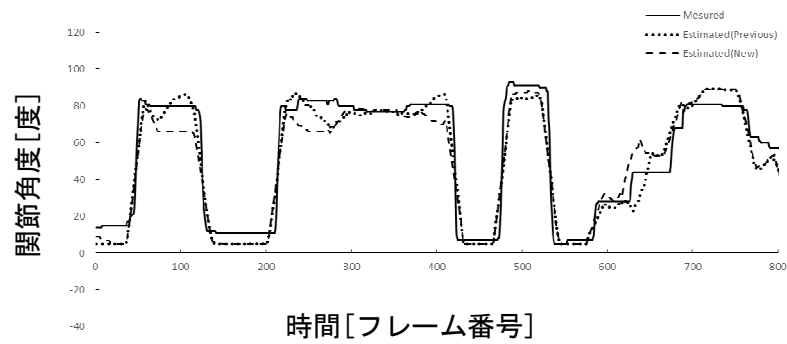
図 5.7 CG による推定結果例（各写真の右が入力の手指形状，左が推定結果の CG）



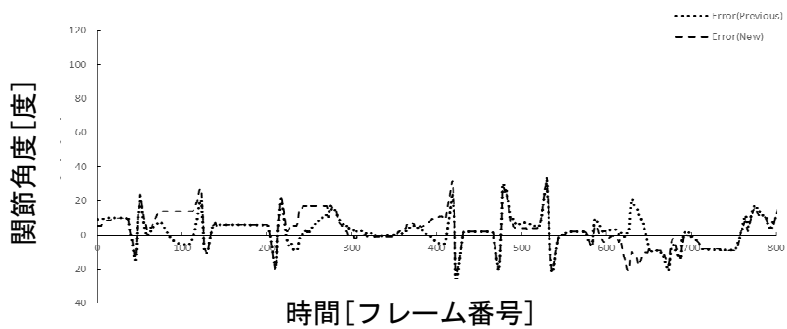
(a-1) 示指 PIP 関節推定結果



(a-2) 示指 PIP 関節推定誤差



(b-1) 拇指 IP 関節推定結果



(b-2) 拇指 IP 関節推定誤差

図 5.8 推定結果例

5.4 まとめ

本章では、モーションキャプチャシステムのように、広い空間で自由に動き回り、様々な姿勢をとるユーザを対象とする状況に対応できる手指形状推定システムの構築を目指した。しかし、2章と3章で示した方法では、カメラ位置が空間に対して固定されているため、撮像される手指が非常に小さかったり、体で隠されてしまったりするという問題がある。

そこで、超小型カメラと小型コンピュータによる装着型ハードウェアを使った手指形状推定システムによって、オペレータの動きを妨げずに、手指形状推定できるシステムの実現を図った。ハードウェアに使用する超小型カメラは、小さいもので $7\text{mm}\times 7\text{mm}$ 程度と非常に小型であるので、装着してもオペレータへの負担が少なく、自由な動きを妨げない。

ただし、小型コンピュータは通常の PC よりもスペックが低いため、高速に動作する画像認識による手指形状推定アルゴリズムが必要となる。そこで、本研究では、特徴量を次元圧縮することで計算コストの低減させ、低スペックな小型コンピュータ上であっても高速高精度に動作する手指形状推定システムの実現を目指した。従来、画像特徴量では HLAC の 25 パターンを使用していた。しかし、25 パターンの内の 1 パターンが最も重要であることが分かった。そこで、この 1 パターンのみを使うことで、全体の特徴量を 1,600 次元から 64 次元に低次元化した。

評価実験では、小型カメラと小型コンピュータを組み合わせたハードウェアを使って推定実験を行った。小型カメラは、手の甲から約 7cm 程度上の部分となるように装着しており、ユーザの動作への影響を最小限に抑えている。推定精度の評価は、データグローブによる各指関節の計測値との推定誤差で検証している。

評価実験の結果、推定誤差の平均値と標準偏差は -5.7 ± 7.2 度であり、推定速度の平均値は 25.0ms であった。従来手法での推定精度が -6.0 ± 6.9 度で、推定速度の平均値が 158.7ms であったので、ほとんど推定精度を落とすことなく、推定速度を大幅に高速化することに成功している。

一連の結果から、自由に動作するユーザによる手指形状推定システムとして、本手法が有効であることが示唆された。

6 章 結論

本稿では、全人工システムを、自由に動きながら、物理世界と同じ手指動作で操作するための手指形状推定システムの実現を目指した。実現のためには、第1の要素として、過去の推定結果に影響されずに、データグローブのような十分な分解能を持った高速高精度の手指形状推定できる必要があった。また、第2の要素として、把持動作のように、特に自己遮蔽の影響によって推定が難しい4指と拇指の対向動作であっても、高精度に推定できる必要があった。さらに、第3の要素として、モーションキャプチャのようにユーザが自由に動作する状況下であっても、高速高精度に推定できる必要があった。

第2章では、データベース探索範囲を狭めない手法によって、コンピュータグラフィックス制御のための手指形状推定システムの実現を目指した。データベースの探索範囲を狭めない手法として、データベースの探索範囲を狭める必要のない手法として、大まかな形状を表す画像形状比率によるスクリーニングと詳細な形状を表す画像特徴量による類似度計算とする2段階のデータベース探索手法を提案した。この手法であれば、データベースを高速に全探索することができるため、探索空間を狭める検討をする必要がない。データグローブの計測値を真値とした推定精度を検証するために、データグローブを装着した状態で手指形状推定を行い、データグローブの実測値と推定値を同時に取得する評価実験を行った。その結果、推定誤差の平均値 -2.11 度、標準偏差 ± 14.14 度、処理速度 80fps 程度と高速高精度に実現できていることが確認できた。

第3章では、ロボットの把持動作で最も必要となる4指と拇指の対向動作を高精度に推定するための手指形状推定システムの実現を目指した。ロボットによる把持動作を実現するためには、4指と拇指の対向動作を高精度に推定できることが欠かせない。しかし、把持動作は、他の手指形状よりも自己遮蔽の影響の大きく推定が非常に難しい。特に、1台のカメラを用いた手指形状推定システムでは、同じ手指のシルエット画像であっても手指形状が違うために、4指と拇指の対向動作を高精度に推定することは難しい。そこで、本章では、直交関係にある2台のカメラを使い、撮像されている手指面積の大きい方のカメラから4指の手指形状、小さい方のカメラで拇指対向の手指形状を推定するシステムを提案する。この手法により、1台では死角となってしまう部分についても高精度に推定することができる。また、通常、複数のカメラを使う場合には、カメラ同士の厳密な相対位置を検出しておく必要があるが、本手法では厳密な相対位置の検出の必要のない推定手法を使うことで、簡単にシステムを設置することができるシステム構築を実現した。評価実験では、右手にデータグローブを装着した状態で左手で推定を行い、右手と左手を注意深く同期して動かすことで、真値となるデータグローブの実測値と推定値を同時に取得して評価実験を行った。その結果、推定誤差の平均値と標準偏差は示指 PIP 関節で 0.45 ± 14.57 度、拇指 CM 関節で 4.7 ± 10.82 度であり、最も重要となる4指と拇指の対向動作についても正確に推定できていることが確認できた。ま

た、実際にロボット操作へ応用したシステムでは、ロボットを通じた物体把持動作によって、安定して対象の人工物を把持することができた。

第 4 章では、モーションキャプチャシステムのように、広い空間で自由に動きまわり、様々な姿勢をとるユーザを対象とする状況に対応できる手指形状推定システムの構築を目指した。2 章と 3 章で示した方法は、カメラ位置が空間に対して固定されているため、撮像される手指が非常に小さくなったり、体で隠されてしまったりするという問題があった。そこで、超小型カメラと小型コンピュータによる装着型ハードウェアを使った手指形状推定システムによって、オペレータの動きを妨げずに、手指形状推定できるシステムの実現を図った。ハードウェアに使用する超小型カメラは、小さいもので $7\text{mm} \times 7\text{mm}$ 程度と非常に小さいので、装着してもオペレータへの負担が少なく、自由な動きを妨げない。また、試作機では、小型カメラを手の甲から約 7cm 程度上の部分となるように装着しており、ユーザの動作への影響を最小限に抑えている。ただし、小型コンピュータは通常の PC よりもスペックが低いため、高速に動作する手指形状推定手法が必要となる。そこで、本章では、画像特徴量の算出に使用していた HLAC を位置情報のみ使用する方法で、推定精度を落とさずに 1600 次元から 64 次元に次元圧縮する手法を提案した。この手法により、従来と同程度の推定精度を維持しつつ、高速に推定することができるようになった。評価実験は、第 2 の研究と同様に、右手にデータグローブ、左手に小型カメラと小型コンピュータを組み合わせたハードウェアを組み合わせた試作機を装着して実施した。その結果、推定誤差の平均値と標準偏差は -5.7 ± 7.2 度であり、推定速度の平均値は 25.0ms であった。一方、1600 次元の画像特徴量を使う手法では、推定精度が -6.0 ± 6.9 度で、推定速度の平均値が 158.7ms であり、ほとんど推定精度を落とすことなく、推定速度を大幅に高速化することに成功した。

一連の結果は、提案した手指形状推定システムが、最終目標に必要な 3 つの要素を満たしていることを示し、本研究が最終目標とする物理世界と同じ手指動作で、自由に動きながら、コンピュータ・グラフィックス制御とロボット操作を可能にする手指形状推定システムが達成されたと結論付けることができる。

謝辞

本論文を作成するにあたり，日頃より有意義なご指導を与えてくださった，筑波大学システム情報系 星野聖教授に深く感謝の意を表します．また，本論文を査読して下さるとともに，大変貴重なご意見を頂きました筑波大学システム情報系 丸山勉教授，同社会工学系 秋山英三教授，同システム情報系 星野准一准教授，望山洋准教授に心からお礼申し上げます．また，日頃より研究生活を共にし，時に研究の有意義な話を，時に実験の協力を，そして時に日々の生活を楽しくしてくれました医用生体工学研究室の皆様に深く感謝いたします．

参考文献

- [1] I. Oikonomidis, N. Kyriazis, and A. Argyros, "Efficient model-based 3D tracking of hand articulations using Kinect," in BMVC 2011, 2011.
- [2] N. Kyriazis, I. Oikonomidis, and A. Argyros, "A GPU-powered computational framework for efficient 3D model-based vision," ICS-FORTH, TR420, , 2011.
- [3] S.Lu,D.Metaxas, D.Samaras and J.Oliensis: "Using multiple cues for hand tracking and model refinement," Proc. CVPR2003, 2, pp.443-450 (2003)
- [4] Y.Wu, J.Lin and T.S.Huang:"Analyzing and capturing articulated hand motion in image sequences," IEEE Transactions on Pattern Analysis and Machine Intelligence, 27 (2005) 1910-1922.
- [5] 島田伸敬, 今井章博, 久野義徳, 三浦純:"緩やかな制約知識を利用した単眼視動画像からの関節物体の形状と姿勢の同時推定," 信学論, J81-D-II, 1, pp.45-53 (1998)
- [6] 亀田能成, 美濃導彦, 池田克夫:"シルエットを利用した手指の 3 次元形状推定法," Proc.MIRU'92, II, pp.239-246 (1992)
- [7] E.Ueda, Y.Matsumoto, M.Imai,and T.Ogasawara:"Hand pose estimation for vision-based human interface," IEEE Transactions on Industrial Electronics, 50, 4, pp.676-684, 2003.
- [8] T.Gumpp, P.Azad, K.Welke, E.Oztop, R.Dillmann and G.Cheng:"Unconstrained real-time markerless hand tracking for humanoid interaction," Proc. IEEE-RAS International Conference on Humanoid Robots, CD-ROM (2006)
- [9] J.Cui and Z.Sun:"Visual Hand Motion Capture for Guiding a Dexterous Hand," Proc. 6th Int. Conf. On Automatic Face and Gesture Recognition, pp.729-734 (2004)p.1743-1761 (2007)
- [10] J.M.Rehg,and T.Kanade:"Visual tracking of high DOF articulated structures: an application to human hand tracking," European Conf. Computer Vision, pp.35-46, 1994.
- [11] M.H.Jeong,Y.Kuno, N.Shimada,and Y.Shirai:"Recognition of shape-changing hand gestures," IEICE Transactions on Information and Systems, E85-D, pp.1678-1687, 2002.
- [12] Athitsos, Vassilis, and Stan Sclaroff. "Estimating 3D hand pose from a cluttered image." Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on. Vol. 2. IEEE, 2003.
- [13] Ohkawa, Yasuhiro, and Kazuhiro Fukui. "Hand-Shape Recognition Using the Distributions of Multi-Viewpoint Image Sets." IEICE transactions on information and systems 95.6 (2012): 1619-1627.
- [14] M.H.Jeong, Y.Kuno, N.Shimada,and Y.Shirai:"Recognition of two-hand gestures using coupled switching linear model," IEICE Transactions on In-formation and Systems, E86-D, pp.1416-1425, 2003.

- [15] Yasumoto, Katsuya, Jun-ya Mizuno, and Takio Kurita. "Improvement of Generalization Ability of Kernel-Based Fisher Discriminant Analysis for Recognition of Japanese Sign Language Hand Postures," Yubi-Moji, Using K-means Method." MVA. 2002.
- [16] 今井章博, 島田伸敬, 白井良明: "輪郭の変形学習による 3-D 手指姿勢の認識," 信学論, J88-D-II, 8, pp.1643-1651 (2005)
- [17] V.Athitos and S.Scarloff: "An appearance-based framework for 3D hand shape classification and camera viewpoint estimation," Proc. Automatic Face and Gesture Recognition (2002) 40-45.
- [18] Licsár, Attila, and Tamás Szirányi. "User-adaptive hand gesture recognition system with interactive training." Image and Vision Computing 23.12 (2005): 1102-1114.
- [19] K.Hoshino and T.Tanimoto: "Real time search for similar hand images from database for robotic hand control," IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences, E88-A(2005) 2514-2520.
- [20] K.Hoshino, E.Tamaki and T.Tanimoto: "Copycat hand - Robot hand imitating human motions at high speed and with high accuracy," Advanced Robotics, 21, 15, pp.1743-1761 (2007)
- [21] 谷本貴頌, 富田元將, 森大祐, 以後直樹, 星野聖: "多階層データベースを用いた高精度かつ高速の手指形状推定," 電気学会 C 部門誌, 129-C, 5, pp.800-805, 2009
- [22] K.Hoshino and T.Tanimoto: "Real Time Search for Similar Hand Images from Database for Robotic Hand Control," IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences, E88-A, 10, pp.2514-2520 (2005)
- [23] K.Hoshino and T.Tanimoto: "Realtime Hand Posture Estimation with Self-Organizing Map for Stable Robot Control," IEICE Transactions on Information and System
- [24] K.Hoshino and E.Tamaki and T.Tanimoto: "Copycat hand - Robot hand imitating human motions at high speed and with high accuracy. Advanced Robotics," 21, pp.1743-1761, 2007.
- [25] K.Hoshino and T.Tanimoto: "Realtime hand posture estimation with Self-Organizing Map for stable robot control," IEICE Transactions on Information and Systems, E89-D, 6, pp.1813-1819, 2006.
- [26] 富田元將, 星野聖: "前腕を含む手指画像からの3次元手指形状推定", 映像情報メディア学会誌, 63, 6, pp.822-828, 2009
- [27] N. Otsu and T. Kurita: A new scheme for practical, flexible and intelligent vision systems, Proc. IAPR. Workshop on Computer Vision, pp.431-435 (1998)
- [28] M.Tomida and K.Hoshino: "Visual-servoing control of robot hand with estimation of full articulation of human hand," Key Engineering Materials, 625, pp.728-735, 2014
- [29] K.Hoshino, T.Kasahara, M.Tomida, and T.Tanimoto: "Gesture-World Environment

Technology for Mobile Manipulation – Remote Control System of a Robot with Hand Pose Estimation –,” Journal of Robotics and Mechatronics, 24, 1, pp.180–190, 2012

- [30] M.Tomida and K.Hoshino: “Wearable Device for High-Speed Hand Pose Estimation with a Ultrasmall Camera,” Journal of Robotics and Mechatronics, 27, 2, pp.167–173, 2015
- [31] D.Kim, et al. “Digits: freehand 3D interactions anywhere using a wrist-worn gloveless sensor.” Proceedings of the 25th annual ACM symposium on User interface software and technology. ACM, 2012.

研究業績

○ 査読付き論文誌

- [1] M.Tomida and K.Hoshino: “Wearable Device for High-Speed Hand Pose Estimation with a Ultrasmall Camera,” Journal of Robotics and Mechatronics, 27, 2, pp.167–173, 2015
- [2] M.Tomida and K.Hoshino: “Visual-servoing control of robot hand with estimation of full articulation of human hand,” Key Engineering Materials, 625, pp.728–735, 2014
- [3] K.Hoshino, T.Kasahara, M.Tomida, and T.Tanimoto: “Gesture-World Environment Technology for Mobile Manipulation – Remote Control System of a Robot with Hand Pose Estimation –,” Journal of Robotics and Mechatronics, 24, 1, pp.180–190, 2012
- [4] 富田元將, 星野聖: “前腕を含む手指画像からの3次元手指形状推定”, 映像情報メディア学会誌, 63, 6, pp.822–828, 2009

○ 査読付き国際会議

- [5] M.Tomida and K.Hoshino: “Development of wearable hand motion capture device in immersive VR system,” Proceedings of The Thirty-Eighth International MATADOR Conference, pp.501–506, 2015
- [6] M.Tomida and K.Hoshino: “Development of Wearable Hand Capture Detecting Hand Posture in Immersive VR system,” Proceedings of the 2014 IEEE/SICE International Symposium on System Integration, pp.163–168, 2014
- [7] M.Tomida and K.Hoshino: “Development of Wearable Hand Capture Device Capable of Achieving Hand Posture in Immersive VR System,” Asia-Pacific Conference on Life Science and Engineering(APCLSE), pp.162–171, 2014
- [8] M.Tomida and K.Hoshino: “3D hand posture estimation with single camera by two-stage searches from database,” 3rd International Universal Communication Symposium(IUCS2009), pp.100–106, 2009